

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 11:20:07 : Search time 6876 Seconds  
(without alignments)  
16902.887 Million cell updates/sec

Title: u66617

Perfect score: 2841  
Sequence: 1 GAATCCCGCCTATCCCATAG.....TGGGCAACAAGGGGATTCC 2841

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

Genemdb1:  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mv:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2841	100.0	2841	9	HSU66617
2	2568.4	90.4	3335	9	BC009368
3	2427.2	85.4	2746	9	AF109733
4	1330.4	46.8	3103	6	AX305314
5	1330.4	46.8	3103	10	MMU66620
6	1319.4	46.4	3101	10	MUS15DKZL
7	1308.6	46.1	3086	10	BC026783
8	1246.4	43.9	164813	9	AC025154
9	773	27.2	2557	5	BC049347
10	698.4	24.6	2747	5	BC049347
11	537.6	18.9	1413	12	BT007694
12	537.6	18.9	1698	9	BC002628
13	530.8	18.7	1790	10	BC013122
14	522.8	18.4	1724	9	HSU66619
15	488.8	17.2	606	6	AX395796
16	430	15.1	2379	10	BC005732
17	429	15.1	2297	10	AB003504
18	414.2	14.6	2018	3	AF071503
19	414.2	14.6	2026	3	AY069383
20	414.2	14.6	33444	2	AC014246
21	414.2	14.6	171751	3	AC023697
22	414.2	14.6	174157	3	AC023711
23	414.2	14.6	332029	3	AE003491
24	412.4	14.5	2041	9	HSU66618
25	394	13.9	1674	3	AK112677
26	323.8	11.4	137956	2	AC010918
27	294	10.3	153863	2	AC025361
28	288.2	10.1	110000	2	AC117865_1
29	288.2	10.1	223435	2	AC129640
30	284.2	10.1	237661	2	AC114454
31	284.8	10.0	195301	2	AC134458
32	238.6	8.4	227255	2	AC139317
33	230	8.1	44171	3	AF000265
34	230	8.1	299719	2	AC006858
35	211.8	7.5	294136	2	AC006901
36	203.2	7.2	1060	10	BC012698
37	191.4	6.7	221111	2	AC116568
38	188	6.6	5938	6	AX344810
39	177.4	6.2	236045	2	AC114659
40	177.4	6.2	260277	2	AC114063
41	176.2	6.2	27832	3	CEZK1128
42	168.8	5.9	4006	9	HUMGP1A
43	145	5.1	5938	6	AX344811
44	123.4	4.3	310	6	AX098486
45	117.4	4.1	160431	8	OSJN00176

# ALIGNMENTS

RESULT 1  
HSU66617  
LOCUS  
DEFINITION Human SWI/SNF complex 60 kDa subunit (BAF60a) mRNA, alternatively spliced, complete cds.  
ACCESSION U66617  
VERSION U66617.1 GI:1549242  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 2841)  
Wang, W., Cole, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R., Murchart, C., Kalpana, G.V., Goff, S.P., Yaniv, M., Workman, J.L. and



QY 1501 CCGATCTTCTACTCCAGGTGACAGAGACGACAGAATTAGACAAGCCCTGGGAAT 1560  
 DB 1501 CCGATCTTCTACTCCAGGTGACAGAGACGACAGAATTAGACAAGCCCTGGGAAT 1560  
 QY 1561 CCGGAATACATAGAGGCTCTCCACAGCCCTGATTCGACCAATTTGATTTGGG 1620  
 DB 1561 CCGGAATACATAGAGGCTCTCCACAGCCCTGATTCGACCAATTTGATTTGGG 1620  
 QY 1621 CCTGTGCTGCTGCTCTCATAGTATGTGCTTGTGTTGGGGCGTTCCAGGGGATG 1680  
 DB 1621 CCTGTGCTGCTGCTCTCATAGTATGTGCTTGTGTTGGGGCGTTCCAGGGGATG 1680  
 QY 1681 CTGTGGTTCAGAGCAAGACCAAGATGAAGAGGGTCTCACAAGACCTGTTATCTCT 1740  
 DB 1681 CTGTGGTTCAGAGCAAGACCAAGATGAAGAGGGTCTCACAAGACCTGTTATCTCT 1740  
 QY 1741 TCTTTCACCCCTATCTCTTCCACACCCCAAGCTTCCCTTTCGCCACAAAGTTCCATGTGC 1800  
 DB 1741 TCTTTCACCCCTATCTCTTCCACACCCCAAGCTTCCCTTTCGCCACAAAGTTCCATGTGC 1800  
 QY 1801 CTGTACCTTCCCTGCTTACATAGGACCTTAGATAGTGTGAAGAGAAATGTAGT 1860  
 DB 1801 CTGTACCTTCCCTGCTTACATAGGACCTTAGATAGTGTGAAGAGAAATGTAGT 1860  
 QY 1861 GGTATAGAGTGTGGAAATGAGATGGCTTCAGGCGCAGGTGTCTTCAAGGGACCACTA 1920  
 DB 1861 GGTATAGAGTGTGGAAATGAGATGGCTTCAGGCGCAGGTGTCTTCAAGGGACCACTA 1920  
 QY 1921 ACTGATTCCTACCTTACAGAGACCCAGAGATGGGTTCCCTCTTCCAAAGTCCAGGC 1980  
 DB 1921 ACTGATTCCTACCTTACAGAGACCCAGAGATGGGTTCCCTCTTCCAAAGTCCAGGC 1980  
 QY 1981 CTGTGGGCACTCTATAGCTAGTGTGCTCTCTGATPAACGAATTCATTTCT 2040  
 DB 1981 CTGTGGGCACTCTATAGCTAGTGTGCTCTCTGATPAACGAATTCATTTCT 2040  
 QY 2041 TCCCTTCCCTCCACAGGTTGGAAACAACTCTCCCTCACTTGTGCTCTGATGACTATA 2100  
 DB 2041 TCCCTTCCCTCCACAGGTTGGAAACAACTCTCCCTCACTTGTGCTCTGATGACTATA 2100  
 QY 2101 GAAACCTGCTTGGGCTCCACTGAGCCCAAGTGTAGTCCCAAGCTCTGGGTTGGCT 2160  
 DB 2101 GAAACCTGCTTGGGCTCCACTGAGCCCAAGTGTAGTCCCAAGCTCTGGGTTGGCT 2160  
 QY 2161 GCTGTAGTGTCTTCTCACTCTTATAGTTGGGGTCCACATAGATTTGATTTCT 2220  
 DB 2161 GCTGTAGTGTCTTCTCACTCTTATAGTTGGGGTCCACATAGATTTGATTTCT 2220  
 QY 2221 TTAATGCTCCCTCCAGACACTCTCTGTGGCTGCTTGTGATTCCTCAATCTGCCC 2280  
 DB 2221 TTAATGCTCCCTCCAGACACTCTCTGTGGCTGCTTGTGATTCCTCAATCTGCCC 2280  
 QY 2281 TAAATCCCGGGCATTTGGGTTGGGGGAATCTTGCTTCCCTTCAGAGCCCAAGGATCTC 2340  
 DB 2281 TAAATCCCGGGCATTTGGGTTGGGGGAATCTTGCTTCCCTTCAGAGCCCAAGGATCTC 2340  
 QY 2341 ATCTGGGGAATCTGATTCGACAGAGAGGCTGTCTTCCCTGACAGTTTGGAGATGTACT 2400  
 DB 2341 ATCTGGGGAATCTGATTCGACAGAGAGGCTGTCTTCCCTGACAGTTTGGAGATGTACT 2400  
 QY 2401 CATTCATCTCACTCACTCCAGCTGCTTGCATCTCTTAATGAGAAACGGGCTTAA 2460  
 DB 2401 CATTCATCTCACTCACTCCAGCTGCTTGCATCTCTTAATGAGAAACGGGCTTAA 2460  
 QY 2461 CCAAAAGGGGTAAAAAAGCCCTGGGCAATCTGTCTTCTGCTGCTGCTGCTGCCAGTTG 2520  
 DB 2461 CCAAAAGGGGTAAAAAAGCCCTGGGCAATCTGTCTTCTGCTGCTGCTGCTGCCAGTTG 2520  
 QY 2521 AACACTACTGTGACTTCTTAGGAGCACTGAGAGTGAAGAGGCTGAGAGATATAC 2580  
 DB 2521 AACACTACTGTGACTTCTTAGGAGCACTGAGAGTGAAGAGGCTGAGAGATATAC 2580

QY 2581 GCTGAGTTGGGTTTGTGACTCTTCCCTCTCCCTGCTCAGAGATTTGACTCCCGAGCC 2640  
 DB 2581 GCTGAGTTGGGTTTGTGACTCTTCCCTCTCCCTGCTCAGAGATTTGACTCCCGAGCC 2640  
 QY 2641 CCTGCCCTCAAGCTTTCAGAGCCCTCAGAGTACAGAGACCTTGTGATCTTGGCCCTT 2700  
 DB 2641 CCTGCCCTCAAGCTTTCAGAGCCCTCAGAGTACAGAGACCTTGTGATCTTGGCCCTT 2700  
 QY 2701 GGATCTGAGATGTTTGTGACTCTTCCAGAGAGACCTTCATCTTCTCCAGTTGTA 2760  
 DB 2701 GGATCTGAGATGTTTGTGACTCTTCCAGAGAGACCTTCATCTTCTCCAGTTGTA 2760  
 QY 2761 TCACCCCGAGTTTACATATCCAGAGCTGCAGACTCAACAGAGAGGTTGGAGACAG 2820  
 DB 2761 TCACCCCGAGTTTACATATCCAGAGCTGCAGACTCAACAGAGAGGTTGGAGACAG 2820  
 QY 2821 CTGGGCACAAGGGGGATTC 2841  
 DB 2821 CTGGGCACAAGGGGGATTC 2841  
 RESULT 2  
 BC009368  
 LOCUS  
 DEFINITION  
 Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1, mRNA (cDNA clone MGC:15280 IMAGE:4110499), complete cds.  
 ACCESSION  
 BC009368.1 GI:14424706  
 VERSION  
 BC009368  
 KEYWORDS  
 MGC.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 1 (bases 1 to 3335)  
 Klausner R.D., Collins E.A., Feingold E.A., Grouse L.H., Derge J.G., Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Altschul S.F., Zeeberg B., Buelow K.H., Schefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F., Diatchenko L., Martins K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshyliki S., Carninci P., Prange C., Raha S.S., Iqbal N.A., Peters G.J., Abramson R.D., Mulhally S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A., Whitting M., Madan A., Young A.C., Shcherchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S., Krzywinski M.I., Skalska U., Smalinski D.E., Scherch A., Schein J.E., Jones S.J. and Marra M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 JOURNAL  
 MEDLINE  
 22388257  
 PUBMED  
 12477932  
 REFERENCE  
 2 (bases 1 to 3335)  
 Strausberg R.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (12-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 REMARK  
 COMMENT  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)

Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mcgen@nih.gov](mailto:nisc.mcgen@nih.gov)  
 Akher N., Ayele K., Beckstrom-Sternberg S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P.,  
 Hansen, N., Ho, S.L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,  
 Maduro Q. L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
 Tsoung, C., Vogt, J.L., Walker, M.D., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H and Green, E.D.

BASE COUNT	746 a	954 c	846 g	789 t
ORIGIN				
Query Match	90.4%;	Score 2568.4;	DB 9;	Length 3335;
Best Local Similarity	95.1%;	Prod. No. 0;		
Matches 2798;	Conservative 0;	Mismatches 11;	Indels 133;	Gaps 9;
Qy	31	CTGAGCTCCCGCTCCGCGCGCGCGCGGGGGAACAGCGCGGCGC	-CGGGGGCGGCTCG	89
Db	10	CTGAGCTCCCGTCCGCGCGCGCGCGGGGAACAGCGCGGCGCTCGCGGGGCGCTCG	69	
Qy	90	GGGGGGGGGGGAGTTCCGGTTTCGGTTCCTTTGTGGGCTGCATCGCGCGCTCGGGGAAG	149	
Db	70	GGGGGGGGGGGAGTTCCGGTTTCGGTTCCTTTGTGGGCTGCATCGCGCGCTCGGGGAAG	129	
Qy	150	ATGGCGGCCCGGGCGGGTTTCCAGTCTGTGGCTCCAAAGCGCGGCGCGCGAGGCTCAGGA	209	
Db	130	ATGGCGGCCCGGGCGGGTTTCCAGTCTGTGGCTCCAAAGCGCGGCGCGCGAGGCTCAGGA	189	
Qy	210	GGGGCGGGCGGCTGTGCTCTTTGGGCC -GGCGGAACCTCGGGGCTCTCTGTGCGAATG	268	
Db	190	GGGGCGGGCGGCTGTGCTCTTTGGGCCCGGGCGGAACCTCGGGGCTCTCTGTGCGAATG	249	
Qy	269	GGGCCGCGCTCCGGGTCAAGGGCTGTACCGTCCCCGATGCCGCGAGCGGCTATCCGAGA	328	



Oy	2360	CCACGACAGGCTGTTCCTTCCTGACGTTGGAGATGACATCATTCATTCACGATCC	2419
Db	2470	CCAGCAGAGGCTGTTCCTTCCTGCTTTGGAGATGACTCATTC - CATTCACGATCC	2528
Oy	2420	ACCCTGCTCTGCATCCCTTAATGAGAAACGGGCTAAACCAACGGGTAAAAAAACC	2479
Db	2529	ACCCTGCTCTGCATCCCTTAATGAGAAACGGGCTAAACCAACGAGGT - AAAAAGCC	2587
Oy	2480	CTGGGCCATCCCTGTCTTCCTGTCCCTTGTCTGCCAGTTGACACCTACTGTGACTTCT	2539
Db	2588	CTGGGCCATCCCTGTCTTCCTGTCCCTTGTCTGCCAGTTGACACCTACTGTGACTTCT	2647
Oy	2540	AGGCGACTGAGAGATGAAGAAACGGCTTAGCGCTGGAGAAATACCGCTGAGTTGGTTTGAC	2599
Db	2648	AGGCGACTGAGAGATGAAGAAACGGCTTAGCGCTGGAGAAATACCGCTGAGTTGGTTTGAC	2707
Oy	2600	TCTTCCCTCTCCCGACCTCACAGGATGTGAACCTCCAGGCCCTTGAGGCTTAAGGCTTCAG	2659
Db	2708	TCTTCCCTCTCCCGACCTCACAGGATGTGAACCTCCAGGCCCTTGAGGCTTAAGGCTTCAG	2767
Oy	2660	ACCCCTAGGTTAGCAGCAGGACCTTGTGATCTTGGCCCTTGATCTGAGATGAGTTTGG	2719
Db	2768	ACCCCTAGGTTAGCAGCAGGACCTTGTGATCTTGGCCCTTGATCTGAGATGAGTTTGG	2827
Oy	2720	CATCTTTCGAGAGAGGCTCACATCTCTTCCAGGTTGATCAACCCCGAGTTAGCAT	2779
Db	2828	CATCTTTCGAGAGAGGCTCACATCTCTTCCAGGTTGATCAACCCCGAGTTAGCAT	2887
Oy	2780	TCCGAGGCTGGCAGACTCAACACAGCAAGGCTGGAGACAGCTGGGCAACAAAGGGGAAAT	2839
Db	2888	TCCGAGGCTGGCAGACTCAACACAGCAAGGCTGGAGACAGCTGGGCAACAAAGGGGAAAT	2947
Oy	2840	CC 2841	
Db	2948	TC 2949	

RESULT 3  
AF109733

LOCUS  
AF109733

DEFINITION  
Homo sapiens SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin D1 (SMARCD1) mRNA, complete cds.

ACCESSION  
AF109733

VERSION  
AF109733.1

KEYWORDS  
GI:4566529

SOURCE  
Homo sapiens (human)

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R.

TITLE  
Diversity and specialization of mammalian SWI/SNF complexes

JOURNAL  
Genes Dev. 10 (17), 2117-2130 (1996)

MEDLINE  
96397413

REFERENCE  
AUTHORS  
2 (bases 1 to 2746)

TITLE  
Ring, H.Z., Vamghi, Meyers, V., Wang, W., Crabtree, G.R. and Francke, U.

JOURNAL  
Five SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin (SMARC) genes are dispersed in the human genome

MEDLINE  
Genomics 51 (1), 140-143 (1998)

REFERENCE  
AUTHORS  
3 (bases 1 to 2746)

TITLE  
Ring, H.Z. and Francke, U.

JOURNAL  
Direct Submission

MEDLINE  
Submitted (25-NOV-1998)

REFERENCE  
AUTHORS  
School of Medicine, B205 Beckman Center, Stanford, CA 94305, USA

TITLE  
Location/Qualifiers

JOURNAL  
1. 2746

MEDLINE  
/organism="Homo sapiens"

REFERENCE  
AUTHORS  
/mol\_type="mRNA"

TITLE  
/cd\_xref="taxon:9606"

JOURNAL  
/chromosome="12"

FEATURES  
source



QY 1748 CCTATCTCTTCCACCCAGCTTCCCTTGGCCCAACAAAGTTCCTGCTAC 1807  
 DB 1681 CCTATCTCTTCCACCCAGCTTCCCTTGGCCCAACAAAGTTCCTGCTAC 1740  
 QY 1808 CTCCTCTGCTACATAGAGACCTCTAGATGTTAGAGAGAAATGCTAGTAA 1867  
 DB 1741 CTCCTCTGCTACATAGAGACCTCTAGATGTTAGAGAGAAATGCTAGTAA 1800  
 QY 1868 AGTCTTGGATGATGAGTGGCTCAGAGGAGTGTCTTCAAGGAGACCACTGATC 1927  
 DB 1801 AGTCTTGGATGATGAGTGGCTCAGAGGAGTGTCTTCAAGGAGACCACTGATC 1860  
 QY 1928 CTACCTTCAGAGACCCAGAGTGGG--TTTGGCTCTTCCAGAGCTGAGGCTGT 1984  
 DB 1861 CTACCTTCAGAGACCCAGAGTGGGAGCTTGGCTCTTCCAGAGCTGAGGCTGT 1920  
 QY 1985 GGGCACTCTATAGCTAGTGTGATCTTGGCTCTCTGATTAAGAAATCCATTTCTCT 2044  
 DB 1921 GGGCACTCTATAGCTAGTGTGATCTTGGCTCTCTGATTAAGAAATCCATTTCTCT 1980  
 QY 2045 TCCCTTCACAGGTTTGGAAACAACTCTCTTCACTTGTGGCTGTAGACTACAGAA 2104  
 DB 1981 TCCCTTCACAGGTTTGGAAACAACTCTCTTCACTTGTGGCTGTAGACTACAGAA 2040  
 QY 2105 CCGGTGTTCTGGCTCCAGTCCAGGCTGAGTCCAGGCTGAGGCTGAGGCTGCTG 2164  
 DB 2041 CCGGTGTTCTGGCTCCAGTCCAGGCTGAGTCCAGGCTGAGGCTGAGGCTGCTG 2100  
 QY 2165 TCAGTCTCTCTCACTCTCTTCTAGTGGGCTCCACATCAGTATGAGTGTGCTTTAT 2224  
 DB 2101 TCAGTCTCTCTCACTCTCTTCTAGTGGGCTCCACATCAGTATGAGTGTGCTTTAT 2160  
 QY 2225 TGCTCTCTCCAGACACTCTCTGCTGCTGCTTGTGATCTCCAGTCTGCTTAT 2284  
 DB 2161 TGCTCTCTCCAGACACTCTCTGCTGCTGCTTGTGATCTCCAGTCTGCTTAT 2220  
 QY 2285 CCGGGGCACTTGGGCTGGGGAATCTTGGCTTCCCTTCAGAGCCCGAGGATCTCATCT 2344  
 DB 2221 CCGGGGCACTTGGGCTGGGGAATCTTGGCTTCCCTTCAGAGCCCGAGGATCTCATCT 2280  
 QY 2345 GGGGAATGCTCATCTTCCAGAGAGGCTGCTCTCTGCTGAGTGTGAGTGTGATCT 2404  
 DB 2281 GGGGAATGCTCATCTTCCAGAGAGGCTGCTCTCTGCTGAGTGTGAGTGTGATCT 2340  
 QY 2405 CCATTCACCTCACTCACTCTGCTCTGCTGCTTCTTATGAGAGAAACGGGCTTAAACCA 2464  
 DB 2341 CCATTCACCTCACTCACTCTGCTCTGCTGCTTCTTATGAGAGAAACGGGCTTAAACCA 2399  
 QY 2465 AGGGTAAAAAGCCCTGGGCACTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCT 2524  
 DB 2400 AGGGTAAAAAGCCCTGGGCACTCTCTGCTCTCTGCTCTCTGCTCTCTGCTCTCTGCT 2458  
 QY 2525 CTACTGTGACTCTTCTAGGCACTGAGAGTGAAGAGGCTTAGGGCTGAGAGTACGCTG 2584  
 DB 2459 CTACTGTGACTCTTCTAGGCACTGAGAGTGAAGAGGCTTAGGGCTGAGAGTACGCTG 2518  
 QY 2585 AGTGGGTTTGTACTCTTCTCTCTCTCTCTCTCTCAAGAGTGTGATCTTCCAGGCTCTG 2644  
 DB 2519 AGTGGGTTTGTACTCTTCTCTCTCTCTCTCTCTCAAGAGTGTGATCTTCCAGGCTCTG 2578  
 QY 2645 CCGCTAAAGCTCAAGCCCTCAAGTGAAGAGAGACCTTGTGATCTTGGCCCTTGGAT 2704  
 DB 2579 CCGCTAAAGCTCAAGCCCTCAAGTGAAGAGAGACCTTGTGATCTTGGCCCTTGGAT 2638  
 QY 2705 CTGAGATGATTTTCTCATCTTCTCAAGAGAGCTCACTTCTTCTTCCAGTGTATCAC 2764  
 DB 2639 CTGAGATGATTTTCTCATCTTCTCAAGAGAGCTCACTTCTTCTTCCAGTGTATCAC 2698  
 QY 2765 CCGGAGTTAGCATATCCAGGCTGAGACTCAACAGCAAGGCTG 2812  
 DB 2699 CCGGAGTTAGCATATCCAGGCTGAGACTCAACAGCAAGGCTG 2746

RESULT 4  
 AX305314  
 LOCUS AX305314 3103 bp DNA linear PAT 11-DEC-2001  
 DEFINITION Sequence 65 from Patent WO018188.  
 ACCESSION AX305314  
 VERSION AX305314.1 GI:17644884  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Scurionathia; Muridae; Mus.  
 REFERENCE  
 1 Ishikawa, K., Asai, S., Takahashi, Y., Nagata, T. and Ishii, Y.  
 AUTHORS  
 TITLE  
 JOURNAL  
 METHOD  
 Patent: WO 018188-A 65 22-NOV-2001;  
 School Unidical Person Nihon University (JP)  
 FEATURES  
 source  
 1. 3103  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 BASE COUNT 657 a 925 c 817 g 704 t  
 ORIGIN  
 Query Match 46.8%; Score 1330.4; DB 6; Length 3103;  
 Best Local Similarity 75.3%; Pred. No. 0; Mismatches 431; Indels 265; Gaps 27;  
 Matches 2124; Conservative 0; Mismatches 431; Indels 265; Gaps 27;  
 QY 158 CCGGCGGGTTTCCAGTGTGCTCAAGGCGCGCGCGAGCTTCAAGAGAGCGGG 217  
 DB 1 CCGGCGGGTTTCCAGTGTGCTCAAGGCGCGCGCGAGCTTCAAGAGAGCGGG 60  
 QY 218 GCGGCTCTGCTTGGGCGCGGAACTCGGGGCTCTCTGTGCGAATGGGCGCGCT 277  
 DB 61 GCGGCTCTGCTTGGGCGCGGAACTCGGGGCTCTCTGTGCGAATGGGCGCGCT 120  
 QY 278 CCGGCTCAAGGCTGTACCGCTCCCGGAGGCGGCTTCCAGAGACAGATATG 337  
 DB 121 CCGGCTCAAGGCTGTACCGCTCCCGGAGGCGGCTTCCAGAGACAGATATG 180  
 QY 338 TTGCGAGCAGCCGATGACACTCAGAGGAGCTTCTCATGAGACCCCTGCTATGGGG 397  
 DB 181 TTGCGAGCAGCCGATGACACTCAGAGGAGCTTCTCATGAGACCCCTGCTATGGGG 240  
 QY 398 AACCTTCAGTCCAGCTGCGCTGCGCAAGTATGATCACTCCGCAAGAGACT 457  
 DB 241 AACCTTCAGTCCAGCTGCGCTGCGCAAGTATGATCACTCCGCAAGAGACT 300  
 QY 458 GCCCTCAGCAGATTCAGCAGAGTCCAGAGAGGCGGTCCAAAATCCAAAACCAATGCA 517  
 DB 301 GCCCTCAGCAGATTCAGCAGAGTCCAGAGAGGCGGTCCAAAATCCAAAACCAATGCA 360  
 QY 518 AAGAAAAGAGATGAGCTGACAAATTTCTACCTCAAGAGATTCGTAAGTGGTACAGAA 577  
 DB 361 AAGAAAAGAGATGAGCTGACAAATTTCTACCTCAAGAGATTCGTAAGTGGTACAGAA 420  
 QY 578 TCCAGGCTTATGATGATCTTGGCTTTTGAAGAAATGAGACAGATATCATGAGG 637  
 DB 421 TCCAGGCTTATGATGATCTTGGCTTTTGAAGAAATGAGACAGATATCATGAGG 480  
 QY 638 AAGCGCTAGATATTCAGAGAGGCTTGAAGTCCCATTAAGCAAAAACGAAAGCTGGA 697  
 DB 481 AAGCGCTAGATATTCAGAGAGGCTTGAAGTCCCATTAAGCAAAAACGAAAGCTGGA 540  
 QY 698 ATTTCATTTTCAACACTTTCAATCCGGCTAAGTCAAGTCCGAGATGGGAAAGGAGC 757  
 DB 541 ATTTCATTTTCAACACTTTCAATCCGGCTAAGTCAAGTCCGAGATGGGAAAGGAGC 600  
 QY 758 GTGGCTTCTGGAGCTTGGGTAAGAGAGGCTCTGAGATTCAGCTTGTCCAA 817  
 DB 601 GTGGCTTCTGGAGCTTGGGTAAGAGAGGCTCTGAGATTCAGCTTGTCCAA 660



MMU66620  
LOCUS MMU66620 3103 bp mRNA linear ROD 18-SEP-1996  
DEFINITION Mus musculus SWI/SNF complex 60 kDa subunit (BAF60a) mRNA, complete  
ACCESSION U66620  
VERSION U66620.1 GI:1549248  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 3103)  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R.,  
Murchardt, C., Kalpana, G.V., Goff, S.P., Yaniv, M., Workman, J.L. and  
Crabtree, G.R.  
Purification and biochemical heterogeneity of the mammalian SWI-SNF  
complex  
EMBO J. 15 (1996) In press  
REFERENCE 2 (bases 1 to 3103)  
AUTHORS Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B.R. and Crabtree, G.R.  
TITLE Diversity and specialization of mammalian SWI/SNF complexes  
JOURNAL Genes Dev. 10 (17), 2117-2130 (1996)  
MEDLINE 96397413  
PUBMED 8804307  
REFERENCE 3 (bases 1 to 3103)  
AUTHORS Wang, W., Xue, Y., Zhou, S. and Crabtree, G.R.  
TITLE Direct Submission  
JOURNAL Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford  
University, Beckman Center B207, Stanford, CA 94305-5428, USA  
FEATURES  
source  
1..3103  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/note="similar to mued15k21, GenBank Accession Number  
M25773"  
1..3103  
/gene="BAF60a"  
/locus="1536  
/gene="BAF60a"  
/note="ubiquitously expressed; similar to human BAF60a,  
BAF60b, BAF60c and yeast SWP73 and SWP73b"  
/codon\_start=1  
/product="SWI/SNF complex 60 kDa subunit"  
/protein\_id="AAC52794.1"  
/db\_xref="GI:1549248"  
/translation="MGPAAGGGLYRSPMPGAAYPRPGLPGSRMTPOGSPMPGPGY  
NPSVRPGAAGSGMDOSRRKRPAPQOIQVOQAOVORNRNNAKKKKADILPQRIELV  
PESQAVMDLAEFERLDOTIMKRLDIOELAKRPDKORAKRIFISNPNKASDAED  
GEGTASWELREGLLEDALSKDATTOKRKRSSFRSLVIELDKLXGPDNHLVE  
WHRTATGETDFOVKRPGDVNRCITVLLMDYOPORFLDRRLRLGLIHOTRPVI  
IQALWQYIKTHRLQDPHERFVLCQYLOQIFESQRMFSELPORLHLMLPPEPII  
NHVLSVDNDOKTACDYDIVEVDITLQKMSFLSTASQOEIATLNNKIHETIETI  
NOLKTORSEFMSLPEPSPGINDMLQSCRKLTMTDVGNPEERAEFYRQPMQEA  
VCRVYSVYQGRQRLDQALGIRNT"

BASE COUNT 657 a 925 c 817 g 704 t  
ORIGIN

Query Match 46.8%; Score 1330.4; DB 10; Length 3103;  
Best Local Similarity 75.3%; Pred. No. 0;  
Matches 2124; Conservative 0; Mismatches 431; Indels 265; Gaps 27;

QY 158 CCGGCGGGTTTCAGTCTGTGCTTCAAGCGGCGGCGGAGCCTCAGAGGAGGCGG 217  
DB 1 CCGGCGGGTTTCAGTCTGTGCTTCCGAGCGGTGCGGCGGAGCCTCAGAGGAGGCGG 60  
QY 218 CCGGCGGTGCTGTGCGGCGGCGGAACTCCGCGGCGCTCTCGGAATGGCGCGGCT 277  
DB 61 CGTGCGGCTGTGCGGCGGCGGAACTCCGCGGCGCTCTCGGAATGGCGCGGCT 120  
QY 278 CCGGCTCAAGGCTGTACCGCTCCCGATGCCGAGCGGCGCTATCCGAGACAGATATG 337  
DB 121 CCGGCTCAAGGCTGTACCGCTCTCCGATGCCGCGGCGGCGCTATCCGAGACAGATATG 180

QY 338 TTGCGAGCGAGCCGAATGACACCTCAGGAGCCTTCATGGAGACCCCTGGCTATGGGGG 397  
DB 181 CTGCGAGGCTAGCCGATATACCTCAGGAGCCTTCATGGAGCCTTCGCTATGGGGG 240  
QY 398 AACCTTCAGTCCGACCTGCGCTGAGCCAGTCAAGGATGATCATGTCCTCCGAGAGACT 457  
DB 241 AACCTTCAGTCCGACCTGCGCTGAGCCAGTCAAGGATGATCATGTCCTCCGAGAGACT 300  
QY 458 GCCCTCAGCGATTCAGACAGTCCAGACAGAGCGGTTCMAAATCGAAACACCAATGCA 517  
DB 301 GCACCTCAACGATTCAGACAGTCCAGACAGAGCGGTTCMAAATCGAAATCACATGCA 360  
QY 518 AAGAAAAGAGATGGCTGACAAATTTACCTCAAGAGATTCGAACTGATACAGAA 577  
DB 361 AAGAAAAGAGATGGCTGACAAATTTACCTCAAGAGATTCGAACTGATACAGAA 420  
QY 578 TCCAGGCGCTATATGATCTCTTGCTTTTGAAGAACTGACAGACTATCATGAGG 637  
DB 421 TCAGAGGCTTACATGATCTCTTGCTTTTGAAGAACTGACAGACTATCATGAGG 480  
QY 638 AAGCGCTAGATATTCAGAGAGGCTTGAACGTCCATTAAAGCAAAAACGAGCTGCGA 697  
DB 481 AAGCGCTAGATATTCAGAGAGGCTTGAACGTCCATTAAAGCAAAAACGAGCTGCGA 540  
QY 698 ATTTTCATTTTCAACATTCATCCGCTAAGTCAGATGCGAGAGATGGGGAAGGAGCG 757  
DB 541 ATTTTCATTTTCAACATTCATCCGCTAAGTCAGATGCGAGAGATGGGGAAGGAGCG 600  
QY 758 GTGCGCTCTCGGAGCTTCGCGGTAGAGAGAGCGCTCTCGAGAGATTCAGCTTGTCCAA 817  
DB 601 GTGCGCTCTCGGAGCTTCGCGGTAGAGAGAGCGCTCTCGAGAGATTCAGCTTGTCCAA 660  
QY 818 TATGATGCCACTAAACAAAGAGAGAGTCTCTCTCTTTAACTCTTGTGATGAA 877  
DB 661 TATGATGCCACTAAACAAAGAGAGAGTCTCTCTCTTTAACTCTTGTGATGAA 720  
QY 878 CTGACCAAGACCTGTATAGGCGCAGACCAATCTGAGAGATGAGACAGAGCGGCACT 937  
DB 721 CTGACCAAGACCTGTATAGGCGCAGACCAATCTGAGAGATGAGACAGAGCGGCACT 780  
QY 938 ACCGAGAGACCGATGCGCTTCAGGTGAAGCGGCGAGAGATGTAATGATGATGAT 997  
DB 781 ACCGAGAGACCGATGCGCTTCAGGTGAAGCGGCGAGAGATGTAATGATGATGAT 840  
QY 998 GTCTACTGATGCTGATTTACCAAGCTTCCAGTTTAAATTAGACCCCGCTAGCTTGA 1057  
DB 841 GTCTACTGATGCTGATTTACCAAGCTTCCAGTTTAAATTAGACCCCGCTAGCTTGA 900  
QY 1058 CTCCTGGGATCCATACCCAGACTCTGCTAGTATCATCCAAAGCACTGTGGCAATATAT 1117  
DB 901 CTCCTGGGATCCATACCCAGACTCTGCTAGTATCATCCAAAGCACTGTGGCAATATAT 960  
QY 1118 AAGACACATTAAGCTTCAGAGACCTTCACAGAGGAGGAGTTGTATCTGTGACAAGTACT 1177  
DB 961 AAGACACATTAAGCTTCAGAGACCTTCACAGAGGAGGAGTTGTATCTGTGACAAGTACT 1020  
QY 1178 CAGCAGATCTTTAGACTCAACGATAGAGTTTTCAGAGATCCCTCAGCGGCTCCAGTCC 1237  
DB 1021 CAGCAGATCTTTAGACTCAACGATAGAGTTTTCAGAGATCCCTCAGCGGCTCCAGTCC 1080  
QY 1238 TTGCTTATGCAACCAAGACCTATCATATATATCATATGATCATGATGATGATGATGAT 1297  
DB 1081 TTGCTTATGCAACCAAGACCTATCATATATATCATATGATCATGATGATGATGATGAT 1140  
QY 1298 CAGAAAAGAGAGCTTGTATGACATTTATGAGATTTGAGAGTGAACACTTGAAGCCAG 1357  
DB 1141 CAGAAAAGAGAGCTTGTATGACATTTATGAGATTTGAGAGTGAACACTTGAAGCCAG 1200  
QY 1358 ATGAATTTCTTTCTGCTGTCACTGCGAGCCAAAGAGATTTGATCTTGAACA----- 1413  
DB 1201 ATGAATTTCTTTCTGCTGTCACTGCGAGCCAGAGAGATTTGCACTTGAACA----- 1260



/db xref="GI:576884"  
/translation="PGFPVCSERCSLRSGSCSGPGTROPVVRMPAPRG  
LVRPMGAAYRPGMLPGSRMTPGSPGPNBVRGLASGMDOSRKRPA  
QOIQOYQOQAVQNRHNNAKKKADKILRIRIYELVPEQATMDLAFERKIDOTIMR  
KRLDIOELKRPDKRKLRTISNTFNPAKSDABEDBSTVSWELRVRGLLEBDAL  
SKYDNTKORRKSSEFKSLVILDKDLVSPDNHLEVMHTATQEDVQVRPGDVN  
VACTUMLMDYOPQFQKPLRLALRLIQTORPVIQALMOYIKHKLQDHEBEFV  
LCKYLOJFESORMKFSEI PORLHALLMPPEPIINHYSVPDNOCKKACYDIVE  
VDDTLKTOWNSFLSLTASQOELATLDNKIHETIETINOKTOREPLSLPESGFIND  
WIOSOCRAARR"  
polya\_signal  
3080..3085  
/gene="DISK21"  
BASE COUNT 657 a 924 c 816 g 704 t  
ORIGIN  
Query Match 46.4%; Score 1319.4; DB 10; Length 3101;  
Best Local Similarity 75.3%; Pred. No. 0;  
Matches 2123; Conservative 0; Mismatches 431; Indels 265; Gaps 28;  
QY 158 CCGGGGGGGTTTCAGTCTGTGGCTCCAGCGGCGCGCGAGCTCCAGAGGGGGGG 217  
DB 1 CCGGGGGGGTTTCAGTCTGTGGCTCCAGCGGCTCCGCGGAGCTCCAGAGAGCGGG 60  
QY 218 CCGGGCTGTGCTTGGGCGCGGAGACTCCGGGCGCTCTGTGCGAATGGGCGCGCT 277  
DB 61 CCGGGGGGGTCTGTGGCGCGGAGACTCCGGGCGCTCCGCTGGAATGGGCGCGG 120  
QY 278 CCGGGCTCAAGGCGTGTACCGGCTCCCGATGCGCGGAGCGGCTATCCAGACCAAGTATG 337  
DB 121 CCGGGCTCAAGGCGTGTACCGGCTCCCGATGCGCGGAGCGGCTATCCAGACCAAGTATG 180  
QY 338 TTGGCAGGAGCGAATGACCTCAGGAGACCTTCATGGGAGCCCGCTGATGGGGGG 397  
DB 181 TTGGCAGGAGCGAATGACCTCAGGAGACCTTCATGGGAGCTTCGATGGGGGG 240  
QY 398 AACCTTCAGTCCGACCTGCGCTGCGCCAGTCAAGGATGATCACTCCGCAAGAGACT 457  
DB 241 AACCTTCAGTCCGACCTGCGCTGCGCCAGTCAAGGATGATCACTCCGCAAGAGACT 300  
QY 458 GCCCTTCAGCAGATCCAGAGGTCAGCAGCAGCGCGCTCAAAATGAAACCAATGCA 517  
DB 301 GCACCTTCACAGATCCAGAGGTCAGCAGCAGCGCGCTCAAAATGAAATCAATGCA 360  
QY 518 AAGAAAAGAAAGATGCTGACAAATTTTACCTCAAGAGATTCGTAACTGTGATCCAGAA 577  
DB 361 AAGAAAAGAAAGATGCTGACAAATTTTACCTCAAGAGATTCGTGAACTGTGATCCAGAA 420  
QY 578 TCCAGGCTTATGATCTCTTGGCTTTGAAAGAACTGGAACAGACTATCATGAG 637  
DB 421 TCCAGGCTTATGATCTCTTGGCTTTGAAAGAACTGGAACAGACTATCATGAG 480  
QY 638 AACGGCTGATATCCAGAGGCTTGAAAGCTCCATTAGCAAAACGGAAGCTGCA 697  
DB 481 AACGGCTGATATCCAGAGGCTTGAAAGCTCCATTAGCAAAACGGAAGCTGCA 540  
QY 698 ATTTTCATTTTCAACATTTTCAATCCGCGTAACTGAGTCCGAGATGGGGAAGGAG 757  
DB 541 ATTTTCATTTTCAACATTTTCAATCCGCGTAACTGAGTCCGAGATGGGGAAGGAG 600  
QY 758 GTGGCTTCCTGGAGCTTGGGCTGAGAGAGGCTCTGAGAGATTCAGCTTTGTCAAA 817  
DB 601 GTGGCTTCCTGGAGCTTGGGCTGAGAGAGGCTCTGAGAGAGGCTCTGAGAGAG 660  
QY 818 TATGATGCACTAAACAAAGAGAAATCTCTCTCTTTTAACTGATCTTGTGATGAA 877  
DB 661 TATGATGCACTAAACAAAGAGAAATCTCTCTCTTTTAACTGATCTTGTGATGAA 720  
QY 878 CTGGCAAGAGCTGTATGGGCAAGCAACATCTGTGATGATGAGACAGAGCCGCACT 937  
DB 721 CTGGCAAGAGCTGTATGGGCAAGCAACATCTGTGATGATGAGACAGAGCCGCACT 780  
QY 938 ACCCAGAGACCGATGCTTCCAGGTGAGAGCGCCAGAGATGTGATTAAGGTACT 997

DB 781 ACCCAGAGACCGATGCTTCCAGGTGAGAGCGCCAGAGATGTGATTAAGGTACT 840  
QY 998 GTCCACTAGATGCTGATTTACAGGCTCCCGAGTTTAAATTAGACCCCGCTAGCTGA 1057  
DB 841 GTCCACTAGATGCTGATTTACAGGCTCCCGAGTTTAAATTAGACCCCGCTAGCTCG 900  
QY 1058 CTCTGGGATCATACCAAGACAGCTGTCAGTATCATCAAGCACTGTGGCAATATAT 1117  
DB 901 CTCTGGGATCATACCAAGACAGCTGTCAGTATCATCAAGCACTGTGGCAATATAT 960  
QY 1118 AAGACATATAGCTCCAGAGACCTTCAAGAGCGGAGTTTGCATCTGTGACAACTG 1177  
DB 961 AAGACATATAGCTCCAGAGACCTTCAAGAGCGGAGTTTGCATCTGTGACAACTG 1020  
QY 1178 CAGCAGTCTTTGACATCAAGTATGAAATTTTACAGATTCCTCAGCGGCTCCATGCG 1237  
DB 1021 CAGCAGATCTTTGAAATCTCAGGAGTGAAGTTCTAGAGATCCCTCAGGCGCTCCAGCG 1080  
QY 1238 TTGCTTATGCGCACAGACCTTATCATATTAATCATGTCAGTGTGAGCCGATGAT 1297  
DB 1081 TTGCTTATGCGCACAGACCTTATCATATTAATCATGTCAGTGTGAGCCGATGAT 1140  
QY 1298 CAGAAAAGACAGCTTTGATGACATGATGTTGAAGTGTAGACACCTTGAAAGCCAG 1357  
DB 1141 CAGAAAAGACAGCTTTGATGACATGATGTTGAAGTGTAGACACCTTGAAAGCCAG 1200  
QY 1358 ATGAATTTCTTTCTGCTGCTCCAGTCCAGCCCAAGAGATTTGATCTTACAA-- 1413  
DB 1201 ATGAATTTCTTTCTGCTGCTCCAGTCCAGCCCAAGAGATTTGATCTTACAA-- 1260  
QY 1414 ----- 1413  
DB 1261 ATCCATGAGACGATAGAACATCAACAGCTGAAGACCCAGAGAGTTTCTGTGAGC 1320  
QY 1414 ----- 1418  
DB 1321 TTGGCCGAGCCCTCAGTTTCATCAATGATGTTGGCTTCCAGTCCGAGTCCAGGCTCAAGA 1380  
QY 1419 CAATGATGATGTTGGGTAAACCCAGAGAGAGAGCCGAGCTGATTTACTTCCAGC 1478  
DB 1381 CGATGATGATGTTGGGTAAACCCAGAGAGAGAGCCGAGCTGATTTACTTCCAGC 1439  
QY 1479 CTTGGGCTCAGAGGCTGTGTCGATGATCTTACTTCCAGAGTGCAGAGACGACAG 1538  
DB 1440 CTTGGGCTCAGAGGCTGTGTCGATGATCTTACTTCCAGAGTGCAGAGACGACAG 1499  
QY 1539 AATTAGAGCAAGCCCTGGGAATCCGGAATACATAGAGGCTTCCCAAGCCCTGATTGCA 1598  
DB 1500 AATTAGAGCAAGCCCTGGGAATCCGGAATACATAGAGGCTTCCCAAGCCCTGATTGCA 1555  
QY 1599 CTGACCAATTTCTGATTTGGGCTGTGCTGCTGCTCATAGTATGCTTGGCTT 1658  
DB 1556 CTGACCAATTTCTGATTTGGGCTGTGCTGCTGCTCATAGTATGCTTGGCTT 1611  
QY 1659 GCTTGGGGGCTTCCAGGGAGATGCTGTGTTCAAGACCAAGACGAGATGAAGAGGCT 1718  
DB 1612 GCTTGGGGGCTTCCAGGGAGATGCTGTGTTCAAGACCAAGACGAGATGAAGAGGCT 1667  
QY 1719 CACA-----AGACACTGTATTCCTCTTCTTCAACCTCATCTTCCACCC 1766  
DB 1668 CACATTTCTGCTCACAAGTCACTGTATCCGCTCTGTATCCCAAGCTGTGCCCGCTCC 1727  
QY 1767 CAGCTTCCCTTTGCCCAAAAGTTCCATGCTGCTGATCCCTCCCTGCTTCAATGAG 1826  
DB 1728 GTGCTGTCCCTCCCTCAGCCCAAGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 1787  
QY 1827 ACCTTAGATGATGATGAGA-GAAGACATGATGATGATGATGATGATGATGATGATG 1885  
DB 1788 TTTACTTCCCTTACTTCAATAGAGACCTTGAATGATGATGATGATGATGATGATGATG 1847  
QY 1886 GCTTCAGGACAGTGTGCTTCAAGGAGACAGCTA-ACGTATCTTACTCT--TCAGAGAC 1942  
DB 1848 TCTTAGAGTCAAGGTGTGCTTGAAGAGACAGCTA-ACGTATCTTACTCT--TCAGAGAG 1907





OY	442	GTCCCGGCAAGAGACTGCCCCCTCAGACATCCAGAGGTTCCAGACGAGGGGTCCTCAAA	501
Db	241	GTCCCCGCAAGAGACTGCACTTCAACAAGATCCAGAGGATCCAGACGAGGGGTCCTCAAA	300
OY	502	TCGAACCAACATGCAAGAAAAAAGAAATGCGCTGACAAATTTCTACTCAAGGATTCG	561
Db	301	TCGAATATACATGCAAGAAAAAAGAAATGCGCTGACAAATTTCTACTCAAGGATTCG	360
OY	562	TGAATGCTGATCCAGATCCAGGCTTATATGATCTCTGGCTTTGAAAGAACTGGA	621
Db	361	GGAATCGGTCCAGATATCACAGGCTTACATGATCTCTGGCTTTGAAAGAACTGGA	420
OY	622	CCAGACTATCATGAGAAACGGCTAGATATCCAGAGGCTTTGAAAGCTCCCATTAAGCA	681
Db	421	CCAGACTATTAATGAGAAACGGCTAGATATCCAGAGGCTTTGAAAGCTCCCATTAAGCA	480
OY	682	AAAAACGAGCTGGGAATTTTCATTTCAACATTTCAATCCGGCTAATGCAGATGCCGA	741
Db	481	AAAAACGAGCTGGGAATTTTCATTTCTTAACACGTTTCATTCGGCTAATGCAGATGCCGA	540
OY	742	GGATGGGAAAGGACGATGAGCTTCTCGAGAGCTTCGGTAGAAGACGGCTCTGAGAGA	801
Db	541	GGATGGGAAAGGACGATGAGCTTCTCGAGAGCTTCGGTAGAAGACGGCTCTGAGAGA	600
OY	802	TTCAAGCTTTGTCCTCAATATATGATGCCACTAAACAAAAGAGAAATCTCTCTCTTTAA	861
Db	601	CGCGGCTTTGTCCTCAATATATGAGCCACCAACAAAAGAGAAATCTCTCTCTTTTAA	660
OY	862	GTCCTTGGTGAATGAACCTGGAACAAAGACTGTATATGGCCACACACCATCTGATGAATG	921
Db	661	GTCCTTGGTGAATGAACCTGGAACAAAGACTGTATATGGCCACACACCATCTGATGAATG	720
OY	922	GCACAGAGCCGCCACTACCCAGAGAACCGATGAGCTTCAGGTTGAAGCGCCAGAGAGATG	981
Db	721	GCACAGAGCCGCCACTACCCAGAGAACCGATGAGCTTCAGGTTGAAGCGCCAGAGAGATG	780
OY	982	GAATGTACGGTGTACTGTCTACTGATGCTGGATTACCAAGCTCCGCCATTAAATTGA	1041
Db	781	GAATGTACGGTGTACTGTCTGCTGATGCTGGAATACCAAGCCCCCGAGTTAAATTGA	840
OY	1042	CCCCCGCTAGCTGCACTCTCGGGCATTCATATCCCAAGCTGTCCAGTATCATCAAGC	1101
Db	841	CCCCCGCTAGCTGCGGCTCTTGGGATCATATACAGACAAGTCCAGTATCATCAAGC	900
OY	1102	ACTGTGGCAATATTTAAGACACATTAAGCTCCAGGACCTCCAGAGGGGGATTGTAT	1161
Db	901	ACTGTGGAGATATATTTAAACACACAAAGCTCCAGGACCTCCAGAGGGGGATTGTCT	960
OY	1162	CTGTGACAAGTACCTGACAGACAGATCTTTGAGACTCAAGTATGAAGTTTTCAGAGATCC	1221
Db	961	CTGTGACAAGTACCTGACAGAGATCTTTGAATCTCAGGAGATGAAGTTCTCAGAGATCC	1020
OY	1222	TCAGGCGCTCATGCTCTTGTATTATGCCACCAAGACTATTCATTAATCATGTTCATAG	1281
Db	1021	TCAGGCGCTCCACGCTCTTGTATTATGCCACCAAGGCTCATTCATTCATGTTCATAG	1080
OY	1282	TGTTGACCCGGAATGATCGAAGAAAAGACGCTTTGATATCAATGATGTTGAAGTGAAGA	1341
Db	1081	TGTGTACCCCAATGATCGAAGAAAAGACGCGTGTATATCAATGATGTTGAAGTGAAGA	1140
OY	1342	CACCTTGAAAGACCCGATGAATTTCTTTCTGCTGTCCATCCGACCAACAGAGAGATTCG	1401
Db	1141	CACCTTGAAAGACCCGATGAATCTTTCTTGTGTCCATCCGACCAACAGAGAGATTCG	1200
OY	1402	TACTCTAGACAA-----	1413
Db	1201	CACCTCTAGACAAAGATTCATGAGCATGAGCATCAACCAAGCTGAAGACCAAGCG	1260
OY	1414	-----	1413
Db	1261	AGAGTTCATGTTGAGCTTTGCCCGAGCCCTCAGGGTTTCATCAATGATTTGCGTTCAATC	1320

QY	1414	-----CAAGCAATGACTGATGTGGTGGTAACTCCAGAGAGAGCGCG	1458
Db	1321	CCAGTGCAGGAGCCTCAAGACGATGCTGATGTGTGGTAACTCCGAGAGAGCGCTCG	1380
QY	1459	AGCTGAGTTCTACTTCCAGGCTTGGGCTTCAGAGAGCTGTGTGGCATCTTCTACTCAA	1518
Db	1381	TGCTGAGTTCTACTTCCAGGCTTGGGCTTCAGAGAGCTGTGTGGCATCTTCTACTCAA	1440
QY	1519	GGTGCAGCAGAGCAGCAAGATTAGACAAGCTTGGGAATCCGAAATCATAGGGCCT	1578
Db	1441	GGTGCAGCAGAGGCGCAGAGTTAGACAAGCCTGGGAATCCGAAACATATGGGCCT	1500
QY	1579	CTCCACAGCCCTGATTTGACTGCACCAATTTCTGATTTTGGGCGCTGTGCTCGCTC	1638
Db	1501	CT---GTGGCCCTAGCTGTGCTGCACCGGATTC---CTTGGCGCTGTGCTCGCTC	1553
QY	1639	ATAGTATCTGCGCTTGCTTGCTTGGGCGGCTTCAGAGGAGTGTGTTTCAAGACAA	1698
Db	1554	AGTGTACTGTCTTGCTTGCTTGGGCGGCTTCAGAGGAGTGTGTTTCAAGACAA	1609
QY	1699	GACCAGATGAAGAGGCTCTACA-----AGACACCTGTATCTCTTCTTTC	1746
Db	1610	TGTACATATGAAGAGGCTCTACAATTTCTGTCTCAGAGTCACTGTTATCCGCTCTGA	1669
QY	1747	ACCTTATCTCTTCCACACCCAGCTTCCCTTTGGCCCAAAAGTTCCCATGTGCTGTAC	1806
Db	1670	CCCCAGTGTGCCCCCGCTGCTGTGCCCCCTTCAACCCACCCCGCTCAGTCTCTCC	1729
QY	1807	CCCTCCCTGGCTCATATAGAGACCTCTATAGTGTATAGAGAGAAC-ATGTAGTGTGA	1864
Db	1730	CCATCAGGCTCTGTGTGCTCTTACCTCCCTATCTATATAGAGACCTCTATAGTGTGA	1789
QY	1865	ATGAGTCTTGAATGATTGTGCTTCAGGCGAGTGTCTTTCAGAGGAGCAGCTA-ACT	1923
Db	1790	GAGAACCAAGAGTGGGGGCTCCTGTAGTCAAGTGTCTTTCAGAGGAGCAGCTA-ACT	1849
QY	1924	GATCTTACCT--TCAGAGACCCAGAGATTGGTTTGCCTCTTCTTCAAGACTCAGGC	1981
Db	1850	GATCTCTCCCTGTTCAGAGACCTTACGCTTGGAGACTATCCCTGTCTGAGCCTCAGGCC	1909
QY	1982	TGTGGGCACTTAAACTAGTGTATGTGGCTCTCTGTATACAGAAATCCAAATTCCTT	2041
Db	1910	T-AGGGAGTGTGAAGCTAGCTGACTTGTGCTCCCGTAGCTTGAAC---TTCCTC	1963
QY	2042	CCCTTCCCTCAGAGGTTTGAACAACTCTCCCTTCACTTGTGCTCCCTGTAGACACTACG	2101
Db	1964	CTTCCCTCCGAGGTTGGGGCAGAGG-----CTCTTATCCTCTGGCAGCTAAG	2013
QY	2102	AAACCCGTGTTTGTGCTCACTGAGGCCAGGCTCACTCCCAAGCCTCTGGGTTGGCCTG	2161
Db	2014	GAGC-----CTGGGCTTCACTGAGGCCCGGGTGGTCTCCCTCTGAGCTTAAAC	2066
QY	2162	CTGTCACTGCTCTCTCACTCTTAACTGTGGGGTTCACATCAAGTATTTGA--GTTTGTTC	2219
Db	2067	TGCTGTCTCACTGCTCTCTCACTCTTAACTGTGGGGTTCACATCAAGTATTTGA	2126
QY	2220	TTTATTTGCTCCCTCCAGACACTTCCCTGTGGTGGCTCTTGTATTTCCCTCAAGACTGCG	2279
Db	2127	ATTGTGTCTCCCTCCAGACACTTCCCTGTGGTGGCTCTTGTATTTCCCTCAAGACTGCG	2186
QY	2280	CTAATCCCGGACATTTTGGTGGGGAACTTGTGCTTTCCTTTCAAGGCCCAGGAGTCT	2339
Db	2187	CTTAAACCAAGCTTTTGGTGGGG-ATCTTGTGCTTCTT-----	2225
QY	2340	CATCTGGGAACTGTCAATTGCGACAGAGGCTGTCTTCTCTGCAGTTTGAAGATGTGAC	2399
Db	2236	-----TGTCACTTCCAGCAGAGACTGTCTTCTCCGCTGTATAGAGATGTGCT	2273
QY	2400	TCATTTCACTTCACTCAGCTCAGCCCTGTGATCTTCAATAGAGAAACGGGCTTAA	2459
Db	2274	TGTTT-----ATTCACTTCACTCAGCTCAGCCCTGTGATCTTCAATAGAGAAACGGGCTTAA	2325
QY	2460	ACCAACGAGGTAAAAAGCCCTGGGCAATCCTGTCTTCTGTCCTTGTCTTCCAGATT	2519

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum





through the I.M.A.G.E. Consortium/INL at: <http://image.inl.gov>  
Series: IRAK.Plate: 101 Row: 1 Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis, Similarity but not identity to protein.  
Location/Qualifiers

```

1. 2747
   /organism="Danio rerio"
   /mol_type="mRNA"
   /strain="AB"
   /db_xref="taxon:7955"
   /clone="MGC:55436 IMAGE:2639578"
   /tissue_type="whole body, adult, (one male and one female,
   including unfertilized eggs)"
   /clone_id="Sugano kawakami zebrafish DRA"
   /lab_host="DH10B"
   /note="Vector: pME18S-FL3"
28. 1560
   /codon_start=1
   /product="Unknown (protein for MGC:55436)"
   /protein_id="AAH9347.1"
   /db_xref="GI:29387084"
   /translation="MAARSGFQSTPGGGAGPMGPGPPVAGAGCGMGPGCTPSGMRGPG
GPOHMYRSPBMPGYPGRGMPSPASMTPOGSMGPGYGASPVSPGMYNDPSKRPAP
PNOIOYQOONRQNRKAKKKMADKLLPRLRELVEPSOAYMDLLAFKKLDOTIMRKR
LDIOEARLKRPIKQKRKRIRITSNTPNAPADEBETVASWELRVEGRLLIEDTVAK
YEATQKRKRSSFPSKSLVLELDKDYLPORNDHLEVHEATATQETDGFQVRRPGVGVR
CTVLLMDYQPOPYKLDLPRLARLGLITQRPVILQALMIVYKTKHLDIPHERFINK
DKYLLQOIFETQRMKFSRIPORLALMLPBPPIIINHISVDNPQKTKACDIDEVND
DLTKQWNSFLSLTASQOELAGLDNKIHEITETINOLKTRREFLSFARPDGFINDM
LQSCDRDLKMTDVGNPBEERBRAEFYFGWMAQNAVRYYSKVQQRQRLBQALGIR
NT"

```

BASE COUNT	789	a	624	c	625	g	709	t
ORIGIN								
Query Match	24.6% Score 698.4; DB 5; Length 2747;							
Best Local Similarity	73.5%; Pred. No. 3,7e-169;							
Matches 937; Conservative	0; Mismatches 321; Indels 16; Gaps 3;							
Dy	146	GAA	GAT	G	G	C	G	C
Db	24	GAA	AT	G	G	C	G	C
Dy	206	AG	G	A	G	G	G	G
Db	84	G	G	A	C	C	G	G
Dy	265	A	A	T	G	G	C	C
Db	144	A	A	G	A	T	G	G
Dy	325	G	A	G	A	C	C	G
Db	204	G	A	G	A	C	C	G
Dy	385	T	G	C	T	A	T	G
Db	264	A	G	T	T	A	C	C
Dy	445	C	C	G	A	A	G	A
Db	318	C	C	G	A	A	G	A
Dy	505	A	A	A	C	C	A	A
Db	369	C	A	A	C	C	A	A
Dy	565	A	C	T	G	T	A	C
Db	429	G	C	T	T	A	C	C
Dy	625	G	A	C	T	A	T	A





599 TTGGCTTTTGAAGAAGAACTGGACAGACATATCATAGAGAAAAGGCTAGATATCCAGAG 658  
 337 TTGGCTTTTGAAGAAGAACTGGACATATCATATGCGGAGCGGCTGACATCCAGAG 396  
 659 GCGCTTGAAGAGCGCTTGAAGAAAAGGAGCTGCGAATTTTCAATTTCTTAACCTTTC 718  
 397 GCTTGAAGAGCGGCTTGAAGAAAAGGAGCTGCGAATTTTCAATTTCTTAACCTTTC 456  
 729 AATCCGGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 778  
 457 AATCCGGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 516  
 779 GTAGAAGAGCGCTTGAAGAAAAGGAGCTGCGAATTTTCAATTTCTTAACCTTTC 838  
 517 GTAGAAGAGCGCTTGAAGAAAAGGAGCTGCGAATTTTCAATTTCTTAACCTTTC 555  
 839 AGGAAGTCT 898  
 556 CGGAAGTCT 615  
 859 CGAAGAACCATCT 958  
 616 CCGAACCAACCATCT 675  
 959 CAGTGAAGCGGCGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1018  
 676 CAGTGAAGCGGCGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 735  
 1019 CAGCTTCCCGAGCTTGAATTAAGACCCCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1078  
 736 CAGCTTCCCGAGCTTGAATTAAGACCCCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 795  
 1079 ACTGCTCCAGTATATATCAACAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1138  
 796 ACCCTCTCAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 855  
 1139 CCGTCAAGCGGAGCTTGAATTAAGACCCCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1198  
 856 TCCCTCAAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 915  
 1199 CCGTCAAGCGGAGCTTGAATTAAGACCCCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1258  
 916 CCGTCAAGCGGAGCTTGAATTAAGACCCCGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 975  
 1259 ATCATCATTAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1318  
 976 ATTCATCATTAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1035  
 1319 GACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1378  
 1036 GACATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1095  
 1379 ACTGCGAGCGCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1418  
 1096 ACGGCGCAACAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1135

RESULT 12  
 BC002628 1698 bp mRNA linear PRI 12-JUL-2001  
 LOCUS BC002628  
 DEFINITION Homo sapiens, similar to SWI/SNF related, matrix associated, actin  
 dependent regulator of chromatin, subfamily d, member 3, clone  
 MGC:3653 IMAGE:3608512, mRNA, complete cds.  
 ACCESSION BC002628.1 GI:12803586  
 VERSION BC002628  
 KEYWORDS MGC.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 1698)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission

JOURNAL  
 Submitted (05-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT: MGC help desk  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc\\_mgc@hgti.nih.gov](mailto:nisc_mgc@hgti.nih.gov)  
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
 Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McEllos, J.C.,  
 McDowell, J., Pearson, R., Snyder, B., Stantir, S., Thomas, P.J.,  
 Tjongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,  
 Zhang, L.-H., and Green, E.D.

FEATURES  
 source  
 1..1698  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:3653 IMAGE:3608512"  
 /tissue\_type="uterus, endometrium adenocarcinoma"  
 /clone\_id="NIH MGC 44"  
 /lab\_host="DH10B-R"  
 /note="vector: pOTB7"  
 139..1551  
 /codon\_start=1  
 /product="Similar to SWI/SNF related, matrix associated,  
 actin dependent regulator of chromatin, subfamily d,  
 member 3"  
 /protein\_id="AA02628.1"  
 /db\_xref="GI:12803587"  
 /translation="MTPEGLQHPVTVORPGMPGARMHQAPMGPPSPYMGSPAVR  
 PGLAPAGMEPARKRAAPPQSGAOSQGPVPTAPARSRAKRRKMDKILPORREL  
 VPESQAVMDLAFERKLDOTIMRKRVIDEALRPMOKRLRLIYSNTFPAKPADAE  
 DSDGSIAWSIELRVGSKLLDPSKOKRKSFPKSLVLELDLGLDPNHLVEMHRTPT  
 TQETDGRQVKKRPPDLVSRCTLLMLDVPQPPKDPRLARLLGHTORSRAIVQALMO  
 YVKNRLQDSHDKETINDKRYFOQIFDPRKFSFELPQRLTALLPPIYINHTISV  
 DPSQOKTACYDIDVEVEEPLKQMSGLSTANQOELSDLSKIHETIESINDLKIQ  
 RDPMLSFGRDKEGVODLRLRSQSDRLDKMTVDVNPPEERAEFFHQWWSQEAVSRYF  
 YCKIQRROELQEOSLVVNT"

CDS  
 366 a 535 c 480 g 297 c  
 ORIGIN

Query Match 18.9%; Score 537.6; DB 9; Length 1698;  
 Best Local Similarity 70.2%; Pred. No. 1.4e-127;  
 Matches 786; Conservative 0; Mismatches 289; Indels 45; Gaps 3;

333 CCGAGACAGAGTATGTTCCAGGAGCGGAGTATGACACTTCAGGACCTTCCATGAGACCC 382  
 175 CAGCGCCCCGGAGTGCCTCTGAGGCCCGGATGCCCCACAGGGGGCCCCCATAGGGCCCC 234  
 383 CCGGCGCTTCCCGTATGAGGAGACCTTCAGTCCGACCTGCGCTGCGCCAGTCAAGGATG 436  
 235 CCGGCGCTTCCCGTATGAGGAGACCTTCAGTCCGACCTGCGCTGCGCCAGTCAAGGATG 294  
 437 GATCAGTCCCGCAAGAGACTTGCCTTCAG-----CAGATCCAGAG 478  
 295 GAGCGCGCCCGCAAGAGAGCGCGCGCGCGCGCGAGAGCCAGGACAGAGCCAGGAGGCG 354

QY 479 GTCCAGCAGCAGCGGTCCAAATCGAAACCAATGCAAAAGAAAAAGAGTGCCTGAC 538  
Db 355 CAGCGGTGCCCCACCGCCCCCGCGGAGCGCAGTGCACAGAGGAGGAGTGCCTGAC 414  
QY 539 AAAATTCTACCTCAAGAGATTCGTGAACCTGGTACCAGATCCAGGCTTATGATCTC 598  
Db 415 AAAATCCTCCCTCAAGAGATTCGGAGCTGGTCCCGAGTCCAGGCTTATGATGACCTC 474  
QY 599 TTGGCTTTTGAAGGAACTGACACAGACTATCATGAGAAACGGCTAGATATCCAAGAG 658  
Db 475 TTGGCATTTTGAGAGAACTGATCAACCATCATGCGAAGCGGTGGACATCCAGAG 534  
QY 659 GCCTTGAACGCTCCCATTAAGCAAAACGAGCTGCGAATTTTCAATTTCTTAACATTTTC 718  
Db 535 GCTCTGAAGAGGCCCATGAAGCAAAAGCGGAGCTGCGACTTATATCTCCAACACTTTT 594  
QY 719 AATCGGCTTAAGTCAGATCCGAGATCGGAGAGGAGCGGTGCTTCTGGAGCTTGG 778  
Db 595 AACCTTGGCAAGCCTGATGCTGAGGATTCGAGCGCAGCATTTGCCCTCTGGAGCTACGG 654  
QY 779 GTAGAAGGACGCTCCTTGAGGATTCAGCCTTTGTCCAAATATGATGCGCACTAAACAAAAG 838  
Db 655 GTGGAGGGAAGCTCTGG-----ATGATCCAGCAACAGAG 693  
QY 839 AGGAAGTCTCTCTCTCTTTAAAGTCTTTGATTTGAACCTGGACAAAGACCTGTATGG 898  
Db 694 CGGAAGTCTCTCTCTCTTCAAGAGTTTGGTCATCGAGCTGGACAAAGATCTTTATGGC 753  
QY 899 CAGACACACCATCTGTTAGATGGACAGGACCGGCTACCCAGGAGACCGATGGCTTC 958  
Db 754 CCTGACAAACCATCTGTTGAGTGGCATCGGACACCCAGCACCAGGAGCGAGCGGCTTC 813  
QY 959 CAGGTGAAGCGCGCAGGAGATGTAATGTPACGGTGTACTGCTCTACTGATGCTGATTAC 1018  
Db 814 CAGGTGAAGCGGCTCGGACCTGAGTGTGCGCTGACGCTGCTCTCATGCTGAGCTAC 873  
QY 1019 CAGCCTCCCCAGTTTAAATTAGACCCCGCCTAGCTCGACTCCTGGGCATCCATCCAG 1078  
Db 874 CAGCCTCCCCAGTTTCAAACCTGATCCCCGCTAGCCGCTGCTGGGCTGCACACAG 933  
QY 1079 ACTGTCAGTGTATCATCAAGCAGCTGGCGAATATATTAAGACATAGCTCCAGGAC 1138  
Db 934 AGCCGCTCAGCCATTTGTCAGAGCCCTGTGGCAAGTATGTGAAGACCAACAGGCTCGAGAC 993  
QY 1139 CCTCAGCAGCGGAGTTTGTATCTGTGACAAAGTACCTGCAGCAGATCTTTGAGACTCAA 1198  
Db 994 TCCCATGACAAGAAATACATCAATGCGGACAAAGTATTTCCAGCAGATTTTGTATGCTCC 1053  
QY 1199 CGTATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCTTGTCTTATGCCACCAAGACT 1258  
Db 1054 CGGCTGAAGTTTCTGAGATTTCCCGAGCGCTCACAGCCCTGCTATTTGCCCTGACCCA 1113  
QY 1259 ATCATCATTAATCATGTATCATGTTGACCCGAATGATCAGAAAGACAGCTTGTAT 1318  
Db 1114 ATTGTATCATCAACCATGTATCAGCTGGAACCCCTTCCAGACCAAGAGAGCGGCTGCTAT 1173  
QY 1319 GACATTGATTTGAAGTGTGATGACACCTTTGAAGACCCAGATCAATTTCTTCTGCTGCC 1378  
Db 1174 GACATTGATGAGGTGGAGAGCCATTAAGGGGCGAGATGAGGAGCTTCTCTCTATCC 1233  
QY 1379 ACTGCCAGCAACAGGAGATTTCTACTCTAGACAAAGA 1418  
Db 1234 ACGGCCAACAGCAGGAGATCAGTGTCTGGACAGTAAGA 1273  
RESULT 13  
LOCUS BC013122 1790 bp mRNA linear ROD 16-APR-2003  
DEFINITION Mus musculus SWI/SNF related, matrix associated, actin dependent  
regulator of chromatin, subfamily d, member 3, mRNA (cdna clone  
MGC:8018 IMAGE:3586276), complete cds.  
ACCESSION BC013122  
VERSION BC013122.1 GI:15341882

KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 1790)  
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,J., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,  
Diatchenko,L., Narusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,  
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,  
Villalón,A., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,  
Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E.,  
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
12477932  
2 (bases 1 to 1790)  
AUTHORS Strausberg,R.  
TITLE Direct Submission  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Günaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 10 Row: 1 Column: 23  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 13385379.  
Location/Qualifiers  
1. 1790  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="MGC:8018 IMAGE:3586276"  
/tissue\_type="Mammary tumor. Metallothionien-TGF alpha  
model. 10 month old virgin mouse. Taken by biopsy."  
/clone\_lib="NCI\_CGAP\_Mam1"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. 1790  
/gene="Smardc3"  
/db\_xref="LocusID:66993"  
/db\_xref="MGI:1914243"  
488. .1633  
CDS

Db 1018 CCGCTCAGCAATTTCCAGGCACCTGTGGCAGTATGTAAACCAACGAGCTACAGCACTC 1077  
 Qy 1141 TCAGACGGGAGTTTCTCATCTGTGACAGTACCTGACAGATCTTTAGACTCAACG 1200  
 Db 1078 CCATGACAAAGAGTACATCAATGATCGCAAGATATTTCCAGAGATTTTGTCTGCCCCG 1137  
 Qy 1201 TATGAGGTTTTCAGAGATCCCTCAGCGCGCTTCATGCTTGTTATGCAACAGAACTAT 1260  
 Db 1138 CCTAAAGTTCTGTGATTTCCCCAGCGCCTTCACAGCCCTGTGCTGCCCTTGACCCCAT 1197  
 Qy 1261 CATCATTTATTCATGTCACTAGTGTGTAACCCGAATGATCAGAAAAAGACAGTTGTTATGA 1320

Db 1198 TGATCATCAACGTCAATACGCTGGACCATCAGACCAAGAAAGACAGCGTCTAIGA 1257

Db 1258 CATAGATGTGAGCTGAGGAACTCCCTGAAGAAGCCAGATGATAGCTTCTCTCTCCAC 1317

Db 1318 GGGCCACCCAGCAGAGATCAGTCTCTGAGACGTAAGA 1355

	RESULT	14		
HSU66619				
LOCUS	1724 bp	mRNA	linear	PRI 18-SEP-1996
HSU66619				

ACCESSION	U66619	GI:1549246
VERSION	U66619.1	
KEYWORDS		

ORGANISM	Homio sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	

**AUTHORS**  
Wang, W., Cote, J., Xue, Y., Zhou, S., Khavari, P. A., Biggar, S. R., Murchard, C., Kalpana, G. V., Gott, S. P., Yaniv, M., Workman, J. L. and Crabtree, G. R.

Journal of Complex Variables  
1996, Volume 1, Pages 1 to 1724

TITLE Diversity and specialization of mammalian SWI/SNF complexes  
JOURN. Genes Dev. 10 (17), 2117-2130 (1996)  
MEDLINE 96397413

REFERENCE AUTHORS TITLE	
3 (bases 1 to 1724)	
Wang, W., Xue, Y., Zhou, S. and Crabtree, G. R.	
Direct Submission	

FEATURES	University, Beckman Center B207, Stanford, CA 94305-5438, USA
SOURCE	Location/Qualifiers 1. .1724

```
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/cell_type="Jurkat T cells"
```

```

- /gene="BAFe0C"
CDS 181..1590
- /gene="BAFe0C"

```

expressed in muscle cells; similar to human BAF60a, BAF60b, BAF60c and yeast SWP73 and SWP73b" /codon\_start=1

```

/translation="MTPTGLQHPPTVVRPGMPSGAMRPHQGRPMARAPRTMAAPCD
/protein_id="AAC50697.1"
/db_xref="gi:1549247"

```

PESQAYMDILAEFRKLDQTIMRGVIOEALIKRPMQOKRKLILYISNTFNAKSDMED  
SDGSIASWELRVEGKLUDDPSQOKRREFFSKSLVIELDKOLYGPNNHLVEWHRTPTT



Qy	2650	AAAGCTTCAGACCCCTCAGGTACAGACAGACCTTGATCTTGCCCTTGATCTGAG	2709
Db	171	AAAGCTTCAGACCCCTCAGGTACAGACAGACCTTGATCTTGCCCTTGATCTGAG	112
Qy	2710	ATGGTTTTGCATCTTTCCAGAGAGCTCACAATTCTTCCAGGTTGTATCACCCTCG	2769
Db	111	ATGGTTTTGCATCTTTCCAGAGAGCTCACAATTCTTCCAGGTTGTATCACCCTCG	52
Qy	2770	AGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAGGCTGGGAGACAG	2820
Db	51	AGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAGGCTGGGAGACAG	1

Search completed: October 11, 2003, 14:36:56  
Job time : 6896 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 11:17:57 ; Search time 517 Seconds  
(without alignments)  
14833.856 Million cell updates/sec

Title: U66617  
Perfect score: 2841  
Sequence: 1 GAATTCGGCTATCCCATG.....TCGCCAAGGGGATTTC 2841

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: N.Geneseq\_190un03.\*
- 2: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.\*
- 3: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.\*
- 4: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.\*
- 5: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.\*
- 6: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.\*
- 7: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.\*
- 8: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.\*
- 9: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.\*
- 10: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.\*
- 11: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.\*
- 12: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.\*
- 13: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.\*
- 14: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.\*
- 15: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.\*
- 16: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.\*
- 17: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.\*
- 18: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.\*
- 19: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.\*
- 20: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.\*
- 21: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.\*
- 22: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.\*
- 23: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2001.DAT.\*
- 24: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.\*
- 25: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1457.4	51.3	2017	21	AAF21750 Human breast and o
2	1330.4	46.8	3103	25	AB199242 Mouse ischaemic co
3	537.2	18.9	1452	25	AAD53119 Human BRG1-associa
4	488.8	17.2	606	24	ABK44460 CDNA encoding colo
5	425.8	15.0	2511	24	AB573872 Human CDNA encodin
6	425.4	15.0	3089	22	AAH98221 Human EST-derived
7	425.4	15.0	3272	22	AAH98319 Human EST-derived
8	414.2	14.6	2010	23	ABL02583 Drosophila melanog

9	414.2	14.6	4138	23	ABL02582	Drosophila melanog
10	408.2	14.4	6809	25	AB235936	Human secretory po
11	397	14.0	1970	24	AB054477	Human ovarian anti
12	388.8	13.7	452	21	AAZ80712	Human colon cancer
13	310.2	10.9	690	21	AAZ80561	Human colon cancer
14	298.6	10.5	2581	25	ABX34564	Human mdtl CDNA SE
15	221.6	7.8	755	25	AAD53120	Human BRG1-associa
16	188	6.6	5938	24	ABN80218	Human chemically m
17	145	5.1	5938	24	ABN80219	Human chemically m
18	123.4	4.3	310	22	AAF79889	Nucleotide sequenc
19	113.6	4.0	285	25	ABX47225	Bovine EST associa
20	96.6	3.4	95223	21	AAF22282	BAC containing rep
21	87	3.1	3824	23	ABL03782	Drosophila melanog
22	82.6	2.9	15500	22	ABA15840	Human nervous syst
23	82.6	2.9	15500	22	AA336609	Human cardiovascular
24	82.6	2.9	15500	25	ABZ73784	Secreted protein g
25	82.6	2.9	15500	25	ABT16889	Human secreted pro
26	82.6	2.9	15500	25	ABZ67378	Human secreted pro
27	79.6	2.8	740	24	ABQ65483	Arabidopsis thalia
28	76	2.7	450	24	ABT07117	Human ovarian can
29	75	2.7	450	25	ABX72995	Human ovarian carc
30	75	2.6	8788	22	AAL35930	Human musculoskele
31	75	2.6	8788	22	AA162919	Human genomic DNA
32	72.8	2.6	8788	25	ABX58918	cDNA encoding nove
33	72.8	2.6	520	22	ABA29582	Probe #8048 for ge
34	72.8	2.6	520	22	AAK10561	Human brain expres
35	72.8	2.6	520	22	AA17316	Probe #7249 for ge
36	72.8	2.6	520	22	AA14206	Probe #10892 used
37	72.8	2.6	520	24	ABX10458	Human genome-deriv
38	69.6	2.4	283	25	ABX85074	Corn ear-derived p
39	60.2	2.1	65	24	ABN51678	Mouse spliced tran
40	60	2.1	65	24	ABN3436	Human spliced tran
41	57.8	2.0	335	25	ABT21717	Breast cancer mark
42	54.6	1.9	687	24	ABQ19226	Oligonucleotide fo
43	54.6	1.9	687	24	ABQ19227	Oligonucleotide fo
44	51.6	1.8	1437	25	ABT19018	Apergillus fumiga
45	51.6	1.8	1437	25	ABT20838	Apergillus fumiga

ALIGNMENTS

RESULT 1	
ID	AAF21750 standard; DNA, 2017 BP.
XX	AAF21750;
AC	27-MAR-2001 (first entry)
XX	
DT	Human breast and ovarian cancer associated antigen gene SEQ ID 137.
XX	
DE	
XX	Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW	nocrotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW	antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
KW	antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
KW	Addison's disease; allergy; autoimmune haemolytic anaemia;
KW	multiple myeloma; thyroiditis; diabetes mellitus; Crohn's disease;
KW	cardiovascular disorder; rheumatoid arthritis; ulcerative colitis;
KW	cardiovascular disorder; wound healing; neurological disease; de.
OS	Homo sapiens.
XX	
PN	WO200055173-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US05881.
XX	
PR	12-MAR-1999; 99US-0124270.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	



PI Rosen CA, Ruben SM;  
XX WPI; 2000-611515/58.  
DR P-PSDB; AAB58847.  
XX New human breast and ovarian cancer associated gene sequences and the  
PT polypeptides encoded by these genes, useful in the prevention,  
PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
PT disorders and neurological diseases -  
XX Claim 1; Page 583; 1299pp; English.  
PS Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
XX proteins AAB59111 - AAB59128. The DNA and protein sequences are  
CC associated with breast and ovarian cancer. Included in the invention are  
CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
CC isolation and characterisation of the DNA and protein sequences of the  
CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
CC antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;  
CC antibacterial; antifungal; antiparasitic and cardiant activity. The  
CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
CC particularly breast and ovarian cancer. The nucleic acid sequences,  
CC proteins, agonists and antagonists may also be used in the diagnosis,  
CC prevention and treatment of immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; cardiovascular disorders such as  
CC myocardial ischaemias; wound healing; neurological diseases such as  
CC cerebral anoxia and epilepsy; and infectious diseases.  
XX Sequence 2017 BP; 428 A; 589 C; 455 G; 536 T; 9 other;  
SQ Query Match 51.3%; Score 1457.4; DB 21; Length 2017;  
Best Local Similarity 98.2%; Pred. No. 0;  
Matches 1548; Conservative 8; Mismatches 10; Indels 10; Gaps 8;  
QY 1272 ATGTCATCAGTGTTCACCCGGAATGATCAGAAAAGACAGCTGTTATGACATTGATGTTG 1331  
DB 4 ATGTCATCAGTGTTCACCCGGAATGATCAGAAAAGACAGCTGTTATGACATTGATGTTG 63  
QY 1332 AAGTGGATGACACCTTGAAGCCAGATGAATTTCTTCTGCTGCCACTGCCAGCCAAAC 1391  
DB 64 AAGTGGATGACACCTTGAAGCCAGATGAATTTCTTCTGCTGCCACTGCCAGCCAAAC 123  
QY 1392 AGGAGATTGCTACTCTAGACAAACAGACAAATGACTGATGTTGGTGGTAAACCCAGAGGAG 1451  
DB 124 AGGAGATTGCTACTCTAGACAAACAGACAAATGACTGATGTTGGTGGTAAACCCAGAGGAG 182  
QY 1452 AGCGCCGAGCTGAGTTCTACTTCCAGCCCTGGGCTCAGAGGCTGTGTGCCGATATTCT 1511  
DB 183 AGCGCCGAGCTGAGTTCTACTTCCAGCCCTGGG-KCAGAGGCTGTGTGCCRATATTCT 241  
QY 1512 ACTCCAAGTGCAGCAGACAGACAAAGAAATTAGACAAAGCCCTGGGAATCCGGAAATACAT 1571  
DB 242 ACTCCAAGTGCAGCAGACAGACAAAGAAATTAGACAAAGCCCTGGGAATCCGGNATACAT 301  
QY 1572 AGGGCTCTCCACAGCCCTGATTCGACTGACCAATTTCTGATTTGGGCCCTGTGCTGC 1631  
DB 302 AGGGCTCTCCACAGCCCTGATTCGACTGACCAATTTCTGATTTGGGCCCTGTGCTGC 361  
QY 1632 CTGCTCATAGTATCTGCTTGGTCTTGTGTTGGGCGTTCCAGGGGATGCTGTTGGTTCA 1691  
DB 362 CTGCTCATAGTATCTGCTTGGTCTTGTGTTGGGCGTTCCAGGGGATGCTGTTGGTTCA 421  
QY 1692 AGGCAAGACCAAGATGAAGGGTCTCACAAGACACCTGTTATCTCTTCTTTCCACCT 1751  
DB 422 AGGCAACACCAAGATGAAGGGTCTCACAAGACACCTGTTATCTCTTCTTTCCACCT 481  
QY 1752 ATCTCTTCCACCCCTGCTTCCCTTTGGCCCAACAAGTCCCATGTCCTGTACCTCC 1811  
DB 482 ATCTCTTCCACCCCTGCTTCCCTTTGGCCCAACAAGTCCCATGTCCTGTACCTCC 541

QY 1812 CCTGGTCTACATAGGACCTCTAGATAGTGTGTAGAGAGAACATGTAGTGTAAATGAGTG 1871  
DB 542 CCTGGTCTACATAGGACCTCTAGATAGTGTGTAGAGAGAACATGTAGTGTAAATGAGTG 601  
QY 1872 CTTGGAATGATTTGGCTCAGCCAGGTGTCTTCAAGGGGACAGCTAACTACTATCCTA 1930  
DB 602 CTTGGAATGATTTGGGCTCAGCCAGGTGTCTTCAAGGGGACAGCTAACTACTATCCTG 661  
QY 1931 CCCTTCAGAGACCCAGGAGTTGGG---TTTCGGTCTCTTCCCAAGACTCAGGCGCTGTGGG 1987  
DB 662 CCCTTCAGAGACCCAGGAGTTGGGAGCTTTCGGCTCTCTTCCCAAGACTCAGGCGCTGTGGG 721  
QY 1988 CACTCTATAAGCTAGTTGATCTTGGCTCTCTCTGATAACAGAATCCAAATTTCTTCTTCC 2047  
DB 722 CACTCTATAAGCTAGTTGATCTTGGCTCTCTCTGATAACAGAATCCAAATTTCTTCTTCC 781  
QY 2048 CTCACAGGTTTGGAAACAAACTCTCCCTTCACTTGTGGCCCTGTAGCACTACAGAAACCC 2107  
DB 782 CTCACAGGTTTGGAAACAAACTCTCCCTTCACTTGTGGCCCTGTAGCACTACAGAAACCC 841  
QY 2108 TGGTTCCTT-GGCTCCACTGAGCCCGAGGTGAGTCCCGAG-CCTCTGGGTTGGCCTGCTGT 2165  
DB 842 TGGTTCCTTGGGCTCCACTGCGCCCGAGGTGAGTCCCGAGCCCTCTGGGTTGRCCTGCTGT 901  
QY 2166 CAGTCTTCTCTCACTCTTCTAGTTGGGCTCCACATCAGTATTTGGAGTTTGTCTTTATT 2225  
DB 902 CAGTCTTCTCTCACTCTTCTAGTTGGGCTCCACATCAGTATTTGGAGTTTGTCTTTATT 961  
QY 2226 GCTCCCTCCACAGACATCCCTGTGGCTGCGCTTGTGATTTCCCTCAGATTCGCCCTAATC 2285  
DB 962 GCTCCCTCCACAGACATCCCTGTGGCTGCGCTTGTGATTTCCCTCAGATTCGCCCTAATC 1021  
QY 2286 CCGGGCAATTTGGGTTGGGGAATCTTGGCTTTCCCTTTCCAGAGCCCGAGGATCTCATCTG 2345  
DB 1022 CCGGGCAATTTGGGTTGGGGAATCTTGGCTTTCCCTTTCCAGAGCCCGAGGATCTCATCTG 1081  
QY 2346 GGGAACTCTCATTCGCCAGCAGAGGCTGTCTCTTCTGAGTTTGGAGATGTGACTCATTC 2405  
DB 1082 GGGAACTCTCATTCGCCAGCAGAGGCTGTCTCTTCTGCTGTTGGAGATGTGACTCATTC 1140  
QY 2406 CATTCACCTACTCACCCTGCTCTGCTGCTTCTTAAATGGAGAAACCGGCTTAAACCAAA 2465  
DB 1141 CATTCACCTACTCACCCTGCTCTGCTGCTTCTTAAATGGAGAAACCGGCTTAAACCAAA 1200  
QY 2466 CGGTAATAAAGCCCTGGGCCATCCCTGCTCTTCTGCTCCCTGCTCCAGTTGACACC 2525  
DB 1201 CGGTAATAAAGCCCTGGGCCATCCCTGCTCTTCTGCTCCCTGCTCCAGTTGACACC 1259  
QY 2526 TACTGGTGAATTTAGGGCACTGAGGAGTGAAGCGCTAGGGCTGGAGATAGCGCTGA 2585  
DB 1260 TACTGGTGAATTTAGGGCACTGAGGAGTGAAGCGCTAGGGCTGGAGATAGCGCTGA 1319  
QY 2586 GTTGGGTTGTGACTCTTCCCTCTCTGCTCTCAGAGATGTGACTCCCCAGCCCCCTGC 2645  
DB 1320 GTTGGGTTGTGACTCTTCCCTCTCTGCTCTCAGAGATGTGACTCCCCAGCCCCCTGC 1379  
QY 2646 CCTCAAGCTTCAGACCCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGCCCTTGGATC 2705  
DB 1380 CCTCAAGCTTCAGACCCCTCAGGTAGCAGCAGGACCTTGTGATCTTGGCCCTTGGATC 1439  
QY 2706 TGAGATGGTGTTCATCTTTCCAGGAGAGCTCACAATTTCTTCCAGGTTGTATCACC 2765  
DB 1440 TGAGATGGTGTTCATCTTTCCAGGAGAGCTCACAATTTCTTCCAGGTTGTATCACC 1499  
QY 2766 CCCGAGTTAGCATATCCAGGCTCGCAGATCTCAACACAGCAAGGTTGGGAGACAGCTGGG 2825  
DB 1500 CCCGAGTTAGCATATCCAGGCTCGCAGATCTCAACACAGCAAGGTTGGGAGACAGCTGGG 1559  
QY 2826 CACAAAGGGGATTC 2841  
DB 1560 CACAAAGGGGATTC 1575

RESULT 2  
 ID AB199242 standard; cDNA; 3103 BP.  
 AC AB199242;  
 DT 07-MAR-2002 (first entry)  
 XX  
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:65.  
 XX  
 KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;  
 KM vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.  
 XX  
 OS Mus musculus.  
 XX  
 PN W0200188188-A2.  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 18-MAY-2001; 2001MO-JP04192.  
 XX  
 PR 18-MAY-2000; 2000JP-0145977.  
 XX  
 PA (UYN1-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
 XX  
 PI Ishikawa K, Asai S, Takahashi Y, Negata T, Ishii Y;  
 XX  
 DR WPI; 2002-034733/04.  
 DR P-PSDB; ABB57044.  
 XX  
 PT Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or  
 PT by determining the expression profile of a gene group comprising these  
 PT genes -  
 XX  
 PS Claim 2; Page 203-208; 2690pp; English.  
 XX  
 CC The present invention describes a method for examining ischaemic  
 CC conditions, comprising measuring the expression levels of particular  
 CC genes (1) in a test sample or determining the expression profile of a  
 CC gene group in the sample comprising genes selected from (1). The method  
 CC is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring  
 CC expression levels of particular genes (AB199202 to AB199912, encoding  
 CC the protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The  
 CC expression levels or expression profiles produced by these genes are  
 CC used as an indicator when screening for ischaemic condition-improving  
 CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914  
 CC represent PCR primers for a mouse ischaemic condition related sequence,  
 CC which are used in the exemplification of the present invention.  
 CC  
 SQ Sequence 3103 BP; 657 A; 925 C; 817 G; 704 T; 0 other;  
 XX  
 Query Match 46.8%; Score 1330.4; DB 24; Length 3103;  
 Best Local Similarity 75.3%; Pred. No. 0;  
 Matches 214; Conservative 0; Mismatches 431; Indels 265; Gaps 27;  
 QY 158 CCGGCGGGTTTCACTCTGTGCTCAAGCGCGCGCGAGCTTCAGAGAGGCGGG 217  
 DB 1 CCGGGGGGTTTCACTCTGTGCTCGAGCGGTGCCGCGGAGCTTCAGAGAGGCGGG 60  
 QY 218 CCGGCGGTCTGCTTGGGCGCGGCGGAACTCCGGGGCTCTGTGGAATGGCGCGCT 277  
 DB 61 CCGGCGGTCTGCTTGGGCGCGGCGGAACTCCGGGGCTCTGTGGAATGGCGCGCG 120  
 QY 278 CCGGCGTCAAGGCGTGAACCGCTCCCGAGTCCCGGAGCGGCGTATCCGAGACAGGTATG 337  
 DB 121 CCGGCGTCAAGGCGTGAACCGCTTCGATGCCCGGGCGGCGCTATCCGAGACAGGTATG 180  
 QY 338 TTGCGAGGAGCGCAATGACACTTCAGGAGCTTTCATGAGGAGCCCTGGCTATGGGGGG 397  
 DB

DB 181 CTGCCAGGTAGCCGATGACACTTCAGGGACCTTCATGGGACTCTTGCTATGGGGGG 240  
 QY 398 AACCTTCAGTCCGACCTGGCTGGCCAGTCCAGGGATGATCAGTCCCGCAAGACCT 457  
 DB 241 AACCTTCAGTCCGACCTGGCTGGCCAGTCCAGGGATGATCAGTCCCGCAAGACCT 300  
 QY 458 GCCCTTCAGAGATCCAGAGGTCAGAGGCGGTCGAAATCGAAATCGAAATCGAATGCA 517  
 DB 301 GCACCTCAAGATCCAGAGGTCAGAGGCGGTCGAAATCGAAATCGAATCGAATGCA 360  
 QY 518 AAGAAAAAGAGATGCTGACAAATTTCTACTCAAGAGATTGGAATCTGTACAGAA 577  
 DB 361 AAGAAAAAGAGATGCTGACAAATTTCTACTCAAGAGATTGGAATCTGTACAGAA 420  
 QY 578 TCCGAGGCTATATGATCTCTTGCTTTTGAAGGAACTGGACACCTATCTATGAGG 637  
 DB 421 TCACAGGCTTACATGATCTCTTGCTTTTGAAGGAACTGGACACCTATCTATGAGG 480  
 QY 638 AAGCGCTATGATATCCAGAGGCTTGAAACGTCCTATTAAAGCAAAAACGAAAGCTGCGA 697  
 DB 481 AAGCGCTATGATATCCAGAGGCTTGAAACGTCCTATTAAAGCAAAAACGAAAGCTGCGA 540  
 QY 698 ATTTTCATTTCTAACACTTTCATTCGGCTTAAATGATGCGAGATGGGAGAGGAGG 757  
 DB 541 ATTTTCATTTCTAACACTTTCATTCGGCTTAAATGATGCGAGATGGGAGAGGAGG 600  
 QY 758 GTGGCTTCTGGAGACTTCGGGTAAAGAGCGGCTCTCGAGAGATTCAAGCTTGTCCAA 817  
 DB 601 GTGGCTTCTGGAGACTTCGGGTAAAGAGCGGCTCTCGAGAGAGCGGCTTGTCCAA 660  
 QY 818 TATGATGCACTTAAACAAAGAGAGGTTCTCTCTTAAAGTCTTGTGATTTGA 877  
 DB 661 TATGATGCACTTAAACAAAGAGAGGTTCTCTCTTAAAGTCTTGTGATTTGA 720  
 QY 878 CTGACAAAGACTGTATGGGCGAGCAACATCTGTGATGACAGAGCCGCCACT 937  
 DB 721 CTGACAAAGACTGTATGGGCGAGCAACATCTGTGATGACAGAGCCGCCACT 780  
 QY 938 ACCGAGAGACCGATGCTTCCAGGTGAAGCGGCCAGAGATGTAAATGACGTGTACT 997  
 DB 781 ACCGAGAGACCGATGCTTCCAGGTGAAGCGGCCAGAGATGTAAATGACGTGTACT 840  
 QY 998 GTTCTACTGATGCTGATTAACAGGCTCCCGAGTTAAATAGACCCCGCTAGCTGGA 1057  
 DB 841 GTTCTACTGATGCTGATTAACAGGCTCCCGAGTTAAATAGACCCCGCTAGCTGGA 900  
 QY 1058 CTCTGGGATTCATACCAAGCTGTCAGTATCATCAAGCACTGTGGCAATATAT 1117  
 DB 901 CTCTGGGATTCATACCAAGCTGTCAGTATCATCAAGCACTGTGGCAATATAT 960  
 QY 1118 AAGACATTAAGCTTCAGAGACCTTACAGAGGCGGAGTTTGTATCTGTGACAAATCTG 1177  
 DB 961 AAGACATTAAGCTTCAGAGACCTTACAGAGGCGGAGTTTGTATCTGTGACAAATCTG 1020  
 QY 1178 CAGCAGATCTTGAAGCTCAACGTATGAAGTTTTCAGAGATCCCTCAGGGCTCCATGCC 1237  
 DB 1021 CAGCAGATCTTGAAGCTCAACGTATGAAGTTTTCAGAGATCCCTCAGGGCTCCATGCC 1080  
 QY 1238 TTGCTTATGCCACGAACTATCATCTATATCATCTATCATCTAGTTGACCCGAATGAT 1297  
 DB 1081 TTGCTTATGCCACGAACTATCATCTATATCATCTATCATCTAGTTGACCCGAATGAT 1140  
 QY 1298 CAGAAAAAGACGCTTGTATGACATTTATGATTTGAATGATGACACTTGAAGCCGAG 1357  
 DB 1141 CAGAAAAAGACGCTTGTATGACATTTATGATTTGAATGATGACACTTGAAGCCGAG 1200  
 QY 1358 ATGAATCTTTTCTGCTGCTGATGACATGCGGCAAGAGAGATTTGACTCTAGACAA 1413  
 DB 1201 ATGAATCTTTTCTGCTGCTGATGACATGCGGCAAGAGAGATTTGACTCTAGACAA 1260  
 QY 1414 ----- 1413  
 DB 1261 ATGCATGAGATGATGAGACCATCAACGAGTGAAGACCGAGAGATTCTATGTTGAGC 1320

Qy	1414	-----CAAG	1417
Db	1321	TTGCCGAGCCCTCAGGTTTCATCAATGATTGGCTTCAGTCCAGTGCAGGGGCGCTCAAG	1380
Qy	1418	ACAATAGCTGATGTGGGTAAACCAGAGAGGAGCGCAGCTGAGTTCTACTTCCAG	1477
Db	1381	ACGATGACTGATGTGGTGAACCCGGAAAGAGAGCGTGTGTGAGTTCTACTTCCAG	1440
Qy	1478	CCCTGGGCTCAGGAGGCTGTGCGGATACCTTACTCCRAGGTGACAGACAGACACAA	1537
Db	1441	CCCTGGGCTCAGGAGGCTGTGCGGATACCTTACTCCRAGGTGACAGACAGACAGACAA	1500
Qy	1538	GAATTAGAGCAAGCCCTGGGAATCCGGAATACATAGGGCCCTCTCCACAGCCCTGATTGC	1597
Db	1501	GAGTTAGACNAGCCCTGGGATCCGAAACACATAGGGCCCTCTCTGCGCCTACGCCCTG	1556
Qy	1598	ACTGCACCAATCTTTGATTTGGGCCCTGTGTGCTGCCTGCCTCAATAGTATCTTGGTCT	1657
Db	1557	GCTGCACCGATTCTCTTTGGGCCCTGTGTGCTGCCTGCCTCAGTGTACCTGTCTTGGTCT	1612
Qy	1658	TGCTTTGGGCGTTCCAGGGATGTGTGGTTCAAGGACACAGACCAGAAATGAAGAGGTC	1717
Db	1613	TGCTTTGAGGCAATCCAGGGGA-----CTTGGCTTCAGGACAGTGTCAAAATGAAGAGGTC	1668
Qy	1718	TCACA-----AGACACCTGTATTCTCTTTCACCTATCTTCTCCACACC	1765
Db	1669	TCACATTTCTGTCTCAGCAGTCACCTGTATCCCGTCTGTACCCAGTCTGCTCCCGTCC	1728
Qy	1766	CCAGCTTCCCTTTGCCCCACAAAGTTCCCATGTGCTGTACCTTCCCTCGTCTACATAG	1825
Db	1729	CGTCGTGTCCCCCTCACCCACCCCGCTCAGCTCCTCCCATCAGGCTCCTGTGTGC	1788
Qy	1826	GACCTCTAGATAGTTTAGAGA-CAGAACATGTAGTGGTAATGAGTCTTGGAAATGGATT	1884
Db	1789	CTCTACCTCCCTATCTACATAGACCTCTAGATAGTGTTAGAGAACACAGAGTGGGC	1848
Qy	1885	GGCCTCAGGCCAGGTGTCTTCAAGGGACACAGTA-ACTGATCCTACCT--TCAGAGA	1941
Db	1849	CTCTCAGGTCAAGTGTGTCTTGAGGAGACACAGCTACACTGATCCTGCCCTTGTCAAGAG	1908
Qy	1942	CCCAGGAGTTGGTTCGTCTCTTCCAGACTCAGGCCCTGTGGGCACTCTATAAGCTA	2001
Db	1909	ACCTAGCCCTGGGAGCTATCCCTGTCTGAGCCTCAGSCCT-AGGGCAGTCTGTAAAGCTA	1967
Qy	2002	GTTGATCTTGGCTCTCCTGATAACAGAAATCCAATTCCTTCTTCCCTCCACAGGTTTGG	2061
Db	1968	GCTGACCTTGGCCCTCCCGGTAGC-----TTGCACTTCTTCCCTCCCTCCGCAAGTTGGG	2023
Qy	2062	AACAACTCTCCTTCACTTGTTCGCCCTGTAGCACTACAGAAACCTGGTTCTTGGCTCC	2121
Db	2024	-----GCAGAGGCTCTCTTACCTCTGGGAGTAAGAGG-----CTGGGGTTC	2066
Qy	2122	ACTGAGCCCCAGGTCAGTCCCGACCCCTCTGGG--TTGGCCTGTGTGTCAGTGTCTCTCA	2179
Db	2067	ACTGAGCCCCGGTGTGTCCCTTGCCTCTGGACTTAACTGTCTCTCAGTGTCTCTG	2126
Qy	2180	CTCCTTAGTTGGGTCACATFAGTATGGA--GTTTGTGTCTTTATTTGCTTCCCTCCAG	2237
Db	2127	ACCCTTA--GGGTCCATGTCAATTTGGAGTGTGTGTAATTTGTGTCTCTCCCTCCAC	2183
Qy	2238	ACACTCCCTGTGGCTGCCCTTGTGATTCCTCAGATCTGCCCTAATCCCGGCATTTGG	2297
Db	2184	ACACTCCGTAGCGCCCGAGTTTAGGATTTCCCTTACACCTGGCCCTAACCCACGCTTTGG	2243
Qy	2298	GTGGGGAACTTTGCCCTTCCCTTTTCAGAGCCCCAGGGATCTCATCTGGGGAATGTGCAT	2357
Db	2244	GTTGGGG-ACTTGGCTTTCTCT-----TGTCAT	2270
Qy	2358	TGCCAGCAGAGGCTGTCTTCTTCTGCAAGTTTGGAGATGTGACTCAATTTCAATTCACCT	2417
Db	2271	TCCCAAGCAGACTGTTCTTCTCTGCTGTTAGAGAGTGGCTTGTTT-----ATTCACT	2324

XX Example 1; Column 55; 31pp; English.  
PS  
XX  
CC The invention relates to human nuclear receptor cofactor polypeptides.  
CC The invention particularly relates to proliferator-activated receptor  
CC (PPAR)-gamma and retinoid X receptor (RXR) cofactor polypeptides.  
CC Polypeptides of the invention are useful for screening compounds that  
CC modulate the interaction of the nuclear receptor with the nuclear  
CC receptor cofactor. Modulators of the invention are useful in preparing  
CC a pharmaceutical composition for treating and/or preventing diseases  
CC or pathologic conditions associated with cell types that express PPAR  
CC receptors. The pathologic conditions treated include metabolic or cell  
CC proliferative disorders such as diabetes, cardiovascular disorders  
CC (e.g. atherosclerosis), renal diseases, neurodegenerative diseases  
CC (e.g. Parkinson's disease, Alzheimer's disease), inflammatory diseases  
CC (e.g. psoriasis, acne), wounds, osteoporosis, infections (e.g. HIV)  
CC or cancer. The invention is useful in gene therapy and as vaccines.  
CC The present sequence is human BRG1-associated factor (BAF) 60c2 DNA.  
CC This sequence is used in the exemplification of the invention.  
XX  
SQ Sequence 1452 BP; 330 A; 447 C; 424 G; 251 T; 0 other;  
Query Match 18.9%; Score 537.2; DB 25; Length 1452;  
Best Local Similarity 70.2%; Pred. No. 1.3e-137;  
Matches 785; Conservative 0; Mismatches 288; Indels 45; Gaps 3;  
QY 325 GAGACGAGTATGTTGCCAGGACCGGATGACACTCAGGACCTTCATGGAGACCC 384  
DB 78 GGGCCCCGGAGTGCCTCTGAGACCCGGATGCCCCACAGGGGGCCCCCATGGCCCCC 137  
QY 385 TGGC-----TATGGGGGGAACCTTCAATCCGACCTGCGCCCATGAGGATGA 438  
DB 138 GGGCTCCCCGTATACGGAGACCCCGCGCTGGACCCCGCTGGCCCCCGGGCATGA 197  
QY 439 TCAGTCCGCAAGAGACTGCTCCCTGAG-----CAGATCCAGCAGGT 480  
DB 198 GCGCGCCCGCAAGCGAGCGCGCCCGCGGAGCGAGCGACGACGACAGCGAGGCGCA 257  
QY 481 CCAGCAGCAGGCGGTCCAAATGGAACCAATGCAAGAAAGAAAGATGGCTGACAA 540  
DB 258 GCGGTGCGCCAGCGCCCGCGGAGCGGCACTGCAAGAGAGAAATGGCTGACAA 317  
QY 541 AATTCTACCTCAAGAGATTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 600  
DB 318 AATCTCTCCCTCAAGAGATTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 377  
QY 601 GGGTTTGAAGAGAACTGAGACCTATCATGAGAGAAAGGCTATGATATCAAGAGC 660  
DB 378 GGCATTGAGAGAGAACTGAGATCAATCATGAGAGAGGCGGTGATCCAGAGGCG 437  
QY 661 CTTGAAAGCTCCCATTAAGCAAAAGAGAGCTGCAATTTTCACTTTCACTTTCA 720  
DB 438 TCTGAAGAGCGCCATGACCAAAAGCGAAGCTGCTATATCTCCAAACATTTTAA 497  
QY 721 TCCGGCTAAGTATGATGCGGAGAGTGGGAGGAGGAGTGGCTTCTGGAGAGTTGGGT 780  
DB 498 CCTCGGAAGCTGATGCTGAGGATTCGACGCGACACATTTGCTCTGGAGGTAGGGGT 557  
QY 781 AGAAGACGCGCTCTGAGAGATTCAAGCTTGTCCAATATGATGACCACTAAACAAAGAG 840  
DB 558 GGAAGGGAAGCTCTG-----ATGATCCGAGCAACAGAGAGG 596  
QY 841 GAAAGTCTCTCTCTTTTAAGTCTTGTGATGATGATGATGATGATGATGATGATGATG 900  
DB 597 GAAAGTCTCTCTCTTTTCAAGAGTTGTGATGATGATGATGATGATGATGATGATGATG 656  
QY 901 AGACAGCACTGCTGTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 960  
DB 657 TGAACAACACTGCTGTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 716  
QY 961 GGTGAAGCGCGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020  
DB 717 GGTGAAGCGCGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 776

QY 1021 GCGTCCCCAGTTTAATTAGACCCCGGCTAGCTGACTCTGGGATCATACCCAGAC 1080  
DB 777 GCGTCCCCAGTTTAATTAGACCCCGGCTAGCTGACTCTGGGATCATACCCAGAC 836  
QY 1081 TCGTCCAGTATCATCAAGCAGCTGTGCAATATATTAAGACATAAGCTTCAGAGCC 1140  
DB 837 CCGTCCAGTATCATCAAGCAGCTGTGCAATATATTAAGACATAAGCTTCAGAGCTC 896  
QY 1141 TCACAGGCGGAGTTTGTATCTGTGACAGTACTGACAGATCTTTGAGACTCAACG 1200  
DB 897 CCATCAACAAGAAATCATCAATGAGGAGACAGATTTTCAGAGATTTTGTATGTCGCG 956  
QY 1201 TATGAAGTTTTCAGAGATCCCTCAGCGGCTCATGCTTGTATGCCAGCAACTAT 1260  
DB 957 GGTGAAGTTTTCAGAGATCCCTCAGCGGCTCATGCTTGTATGCCAGCAACTAT 1016  
QY 1261 CATCATTAATCATGTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1320  
DB 1017 TGTCAATCAACATGTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1076  
QY 1321 CATTAATGTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380  
DB 1077 CATTAATGTAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1136  
QY 1381 TGCAGCCCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1418  
DB 1137 GGTCAATCAACATGTCATGATGATGATGATGATGATGATGATGATGATGATGATG 1174  
RESULT 4  
ABK44460/C  
ID ABK44460 standard; cDNA; 606 BP.  
XX  
XX ABK44460;  
XX  
XX 05-JUN-2002 (first entry)  
XX  
XX cDNA encoding colon tumour protein, SEQ ID No 11.  
XX  
XX Human; colon tumour; vaccine; colon cancer; immunogenic;  
XX immunotherapy; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200212328-A2.  
XX  
XX 14-FEB-2002.  
XX  
XX 31-JUL-2001; 2001WO-US24218.  
XX  
XX 03-AUG-2000; 2000US-223283P.  
XX 28-MAR-2001; 2001US-279763P.  
XX 29-JUN-2001; 2001US-302051P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX King GE, Meagher MJ, Xu J, Secret H;  
XX  
XX WPI; 2002-241739/29.  
XX  
XX New colon cancer polypeptides and polynucleotides, useful as vaccines,  
XX for diagnosing, preventing, and treating colon cancer, and as markers  
XX for the progression of cancer -  
XX  
XX Claim 1; SEQ ID No 11; 147pp; English.  
XX  
XX The invention relates to polynucleotides encoding colon tumour proteins.  
XX The polynucleotides and encoded polypeptides are useful in pharmaceutical  
XX compositions, such as vaccines, for the diagnosis, prevention, and  
XX treatment of colon cancer. Polynucleotide sequences may be used as  
XX hybridisation probes or primers, and in the design and preparation of  
XX ribozyme molecules for inhibiting expression of tumour polypeptides and



```

Db      267 GCGCGCGGGGCGCGGCGGCGAGTACGAGGACCTGCGCATGTGACACGAGGAAACCGGATGC 326
Qy      357 CACCTCAGGAGACCTTCCATGGAGACCCCTG-----GCTATGGGGGAAACCTTCAG 407
Db      327 CCATGCGCTGCTTTCAGAGTGGGACCCCTGCTGGCTCCCATTTGGTGGACGAGCTCCGC 386
Qy      408 TCGGACCTGCGCGGCGGCGAGTCAAGGATGATGATCCCGCAAGACGCTGCCCTCAGC 467
Db      387 TTGAGACTGAGTCCACCCACCATGATGATTCATTCCGAAACCCCTGCTTGTGCCCC 446
Qy      468 AGATCCAGCAGGTCCAGCAGCAGCGGTCCAAATCGAAACCAATGCAAAAGAAAAAGA 527
Db      447 AG-----GGGAGCCTCCCATGCTGCCCGGCGGGGTTAAAGAGAGAGA 494
Qy      528 AGATGCTGACAAATTTCTACCTCAAGATTCGTGAATCTGTACAGAAATCCAGCCT 587
Db      495 AGATGCGCATTAAGTTTCTACCTCAGCGAATCCGGAGCTTGTTCAGAGTCTCAGGCGT 554
Qy      588 ATATGATCTCTTGGCTTTTGAAGAACTGACACGACTATCATAGAGAAACGGCTAG 647
Db      555 ACATGGATCTTGGCTTTTGAAGAGCTGACACGACCATTTGCTCGCAAGCGGATG 614
Qy      648 ATATCCAGAGGCTTGAACGTCCTTAAAGCAAAACGAAAGCTGCGAATTTTCATT 707
Db      615 AGATCCAGAGGCGCATCAAAAAGCCTCTGACACAAAAGGAAAGCTTCGGATCTACATT 674
Qy      708 CTAAACCTTCAATCCGGCTAAGTCAATGCGAGATGCGAGATGGGAGGAGCGTGTCT 767
Db      675 CCAATAGCTTCAATCCAGCAAGCGGAGGAGCATATGTCAGGAACTCGAGGAGACCCCT 734
Qy      768 GGGAGCTT-----GGTAAAGAGAGCGCTCTCGAGAGATGACCTTGTCCAAA----- 817
Db      735 GGGGAAACCCAGCAGGAGGAGCAAGTGTGCTTCTGGGAACTCCGATGGAAGGAAACCTGC 794
Qy      818 --TATATGCACTAAACAAAAGAGAGATTCCTTCTCTTAAAGTCTTGTGATTTG 875
Db      795 TGATATCTCTAGCAAAACAGAGAGAAATTTCTTCAATCTTTAAAGACCTCGTCAATTG 854
Qy      876 AACTGACAAAGACTGTATGGGCGACACCAACCATTTGTGAATGGCACAGACCGCA 935
Db      855 AGCTGACAAAGAGCTGTAGGGGCTGACAAATCACTGGTGGAGTGGACCGGATCCCA 914
Qy      936 CTATCCAGAGACCGATGTGCTTCCAGTGAAGCGCCAGAGATGTGAATGTACGTGA 995
Db      915 CCACCCAGAGACAGATGTGCTTCCAAATAAAGGGCTGGAGCTTCAACGTCAAGTGA 974
Qy      996 CTGTCTACTGATGCTGGAATTACAGGCTCCCGAGTTTAAATTAGACCCCGGCTAGCTC 1055
Db      975 CCTCTCTGCTCATGCTGATCATCAGCTTCCCAATGACAAATTTGACCCCGATTGGCAA 1034
Qy      1056 GACTCTGGGCAATCATATCCAGACTGTTCAGTATCATCAAGCACTGTGGCAATATA 1115
Db      1035 GGTGTGTGGGAGTGCACACGACAGAGGGCGGCATCATGACGGCCCTGTGGCTTTACA 1094
Qy      1116 TTAAACACATTAAGCTTCAGGACCTTCAGAGCGGAGTTTTCATCTGTGACAATACC 1175
Db      1095 TCAAGACAAACCACTGTCAGAGATGGGACAGAGGAGAGTACATCAACTGCACCGTTAAT 1154
Qy      1176 TGAAGAGATCTTTGAGACTCAACGTAATGAAGTTTTCAGAGATCCCTCAGCGGCTCATG 1235
Db      1155 TCGCGCAGATCTCAGTTGTGCGGACTCCGTTTCTCCAGATTTCCATGAAAGCTGGCAG 1214
Qy      1236 CTTTGTATGACCAAGAACCTATCATCAATTAATCATGTATCATAGTGTGACCCGAATG 1295
Db      1215 GGTGTGCGAGCATCCAGACCCCATTTGTATCAACATGTCATTAATGTGACCTTAAG 1274
Qy      1296 ATAGAAAAAGACAGCTTGTATGACATGTATGTAAGTGAACCTTTAAAGACCC 1355
Db      1275 ACCAGAAAGACAGCGCTGTATGACATGATGTGAGGTGACGACCACTGAAGGCC 1334
Qy      1356 AGATGAATCTTTTGTGCTGTCACTGGCCAGCAAGAGATGTGCTACTGACAAACA 1415

```

```

Db      1335 AAATGACAAATTTTCTGCGCTCTACCAACCAATCAGAGAGATCGCTCCCTTGATGTCA 1394
Qy      1416 AGA 1418
Db      1395 AGA 1397

RESULT 6
AAH98221
ID AAH98221 standard; cDNA; 3089 BP.
XX
AC AAH98221;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST-derived coding sequence SEQ ID NO: 78.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition; ss.
XX
OS Homo sapiens.
XX
PN MO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Weirman T;
DR WPI: 2001-476164/51.
DR P-PSDB; AAM23562.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 1; Page 231-232; 1275bp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a cDNA
CC of the invention.
SQ Sequence 3089 BP; 648 A; 954 C; 946 G; 541 T; 0 other;
XX

Query Match 15.0%; Score 425.4; DB 22; Length 3089;
Best Local Similarity 72.6%; Pred. No. 1.5e-106;
Matches 519; Conservative 0; Mismatches 211; Indels 7; Gaps 2;
Qy      556 GATTGCTGAACGTATACAGAAATCCAGGCTATATGATCTTGGCTTTGAAGGAA 615
Db      496 GATTTGGGAGCTGTGCTCCGAGTCCAGGCTTACATGACCTTGGCTTTGAAGGAA 555
Qy      616 ACTGACACGATATCATGAGAAACGGCTAGATATCCAAGAGCGCTTGAACGTCCTAT 675
Db      556 ACTGATCAACATCATGCGGAGCGGTGACATCCAAGAGGCTCTGAAGAGGCCAT 615
Qy      676 TAAGAAAAAGGAGCTGCGAATTTCTATTTCTAACAATTTCATTCGGGCTAAGTCA 735

```

Db 616 GAAGCAAAAGCGAAGCTCGACTCTATATCTTCAACACTTTTAAACCCCTGGAAGCCCTGA 675  
Qy 736 TGCCGAGATGGGAAGGACGGTGGCTTCTGGGAGCTTCGGGTAGGAAGACGGCTCCT 795  
Db 676 TGCTGAGGATCCGACGCGACATTTGCTCTGGGAGCTACGGGTGGAGGGAAGCTCCT 735  
Qy 796 GGAGGATTACGCTTTGTCTCAATATGATGCCACTAAACAAAAGAGAAAGTTCTCTTCCTT 855  
Db 736 GGATGAT-----GTACGTCCTGGCCGCCAGCCAGCAACAGAAAGCGGAAGTTCTCTCTTT 790  
Qy 856 CTTTAAGTCTTGGTGTATGAATCGGACAAAGACCTGTATGGCCAGACACACCATCTGGT 915  
Db 791 CTTCAAGAGTTTGGTTCATCGAGCTGACAAAGATCTTTATGGCCCTGCAACACCACTCGT 850  
Qy 916 AGAATGGCACAGGACCGCACTTACCAGGACGACCGATGGCTTCCAGGTGAAGCGCCAGG 975  
Db 851 TGAGTGGCATCGGACCAACCGACCCAGGACGACGCGCTTCNAGTTGAACGGCCCTGG 910  
Qy 976 AGATGTAATGTACGGTGTACTTCTCTACTGATGCTGGATTACCAAGCCTCCCAAGTTTAA 1035  
Db 911 GGACCTGAGTGTGCGCTGCACGCTGCTCTCATGCTGGACTACCAAGCCTCCCAAGTTCAA 970  
Qy 1036 ATTAGACCCCGCTAGCTGCACTCTGGGATCATACCCAGACTCTCCAGTGTATCAT 1095  
Db 971 ACTGGATCCCGCTTAGCCCGCTGCTGGGCTGCACACAGAGCCGCTCAGCCATTGT 1030  
Qy 1096 CCAAGCACTGGGCAATATATTAAACACATAAGCTCCAGGACCTTCAAGCGGGAGTT 1155  
Db 1031 CCAGGCCCTGTGGCAATGTATGAAGCAACCAAGGCTGCAGACTCCCATGACAGGAATA 1090  
Qy 1156 TGTCTATCTGTACAAGTACCTGCACGACATCTTTGAGACTCAAGCTATGAAGTTTTCAGA 1215  
Db 1091 CATCAATGGGACAAGTATTTCCACGACAGATTTTGTGTTGCCCGCTGAAGTTTCTGA 1150  
Qy 1216 GATCCTCAGCGGCTCCATGCTTGTGCTTATGCCACAGAACTATCATCATTAATCATGT 1275  
Db 1151 GATTTCCCGAGCGCTTCACAGCCCTGCTATTTGCCCTTGAACCAATTTGTCATCAACCATGT 1210  
Qy 1276 CATCAGTGTGACCCGCAATGATCAAGAAAAGACAG-CTTGTATGACATTGATGTTTCAA 1333  
Db 1211 CATCAGTGTGACCCCTTCAGACCCAGAGAGACGGTCTGCTATGACATTGACGTGAAG 1270  
Qy 1334 GTGGATGACACCTTGAA 1350  
Db 1271 GTGGAGGAGCCATTAA 1287

RESULT 7  
AAH98319  
ID AAH98319 standard; cDNA; 3272 BP.  
XX AC AAH98319;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE Human EST-derived coding sequence SEQ ID NO: 176.  
XX  
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200154477-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02687.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 17-JUL-2000; 2000US-0617746.  
PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;  
XX Cao Y, Drmanac RA, Zhang J, Werfman T;  
XX WPI; 2001-476164/51.  
XX P-PSDB; AAM23660.  
XX Isolated polypeptide for treatment of diseases, diagnostics, raising  
XX antibodies and research use -  
XX Claim 1; Page 298-299; 1275pp; English.  
XX The present invention provides the protein and coding sequences of novel  
XX proteins from a variety of organisms, including human, dog, cat, horse,  
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)  
XX from the organism of interest. They can be used in diagnostics,  
XX forensics, gene mapping, identification of mutations, to assess  
XX biodiversity and for nutritional purposes. The present sequence is a cDNA  
XX of the invention.  
XX  
SQ Sequence 3272 BP; 682 A; 1013 C; 1001 G; 576 T; 0 other;  
Query Match 15.0%; Score 425.4; DB 22; Length 3272;  
Best Local Similarity 72.6%; Pred. No. 1.6e-106;  
Matches 579; Conservative 0; Mismatches 211; Indels 7; Gaps 2;  
Qy 556 GATTCTGTGAATCGTATACCAAGATCCCAAGGCTTATGATCTCTTGGCTTTTGAAGGAA 615  
Db 496 GATTCTGGGAGCTGTCTCCCGAGTCCAGGCTTACATGGACCTCTTGGCAITTTGAGAGAA 555  
Qy 616 ACTGGACACACTATCATGAGGAAACGGCTAGATATCCAGAGGCTTGAACCTCCCAT 675  
Db 556 ACTGGATCAAAACCATCATGCGGAAGCGGTGGACATCCAGGAGGCTCTGAAGAGGCCAT 615  
Qy 676 TAAGCAAAAACGGAAGCTGCGAATTTTCACTTTCAACACTTTCAATCCGGCTAAGTCAGA 735  
Db 616 GAAGCAAAAGGGAAGCTGCGACTTATATCTCCACACTTTTAAACCTTGGGAAGCTGA 675  
Qy 736 TGCCGAGGATGGGAAGGACGGTGGCTTCTGGGAGCTTCGGGTAGGAAGACGGCTCCT 795  
Db 676 TGCTGAGGATTCGACGCGCAGCATTTGCTCTCTGGAGCTACGGGTGAGGGAAGCTCCT 735  
Qy 796 GGAGGATTCAGCTTGTCCAAATATGATGCCACTTAACAAAGAGGAGTTCTCTTCTT 855  
Db 736 GGATGAT-----GTACGTCCTGGCCGCCAGCCAGCAACAGAAAGCGGAAGTTCTCTCTTT 790  
Qy 856 CTTTAAGTCTTGGTGTATGAACCTGGACAAAGACCTGTATGGGCGCAGACCAACCATCTGGT 915  
Db 791 CTTCAAGAGTTTGGTTCATCGAGCTGGAACAAAGATCTTTATGGCCCTGACAAACCATCTGT 850  
Qy 916 AGAATGGCACAGGACCCCACTACCCAGGACGCGATGGCTTCCAGGTGAAGCGCCAGG 975  
Db 851 TGAGTGGCATCGGACACCCAGCACCAGGACGACGCGCTTCCAGGTGAACCGCTGG 910  
Qy 976 AGATGTAATGTACGGTGTACTTCTCTACTGATGCTGGATTACCAAGCCTCCCAAGTTTAA 1035  
Db 911 GGACCTGAGTGTGCGCTGCAAGCTGCTCTCATGCTGGACTACCAAGCCTCCCAAGTTCAA 970  
Qy 1036 ATTAGACCCCGCTAGCTGCACTCTGGGCACTCCATACCCAGACTCTCCAGTGTATCAT 1095  
Db 971 ACTGGATCCCGCTTAGCCCGCTGCTGGGCTGCACACAGAGCCGCTCAGCCATTGT 1030  
Qy 1096 CCAAGCACTGTGGCAATATATTAAACACATAAGCTCCAGGACCTTCAAGCGGGAGTT 1155  
Db 1031 CCAGGCCCTGTGGCAGTATGTGAAGACCAACAGGCTGCAGGACTCCCATGACAGGAATA 1090  
Qy 1156 TGTCTATCTGTACAGTACCTGACGAGATCTTTGAGACTCAAGCTATGAAGTTTTCAGA 1215  
Db 1091 CATCAATGGGACAAGTATTTCCAGCAGATTTTGTGTTGCCCGCTGAAGTTTCTGA 1150



QY	DB
1216	GATTCCTCAGGGGCTCATTGCCTTGGCTTAATGCACACAGACCTTATCATCATTAATCATGT
1151	GATTCCTCAGGGGCTCATTGCCTTGGCTTAATGCACACAGACCTTATCATCATTAATCATGT
1276	CATCAGTGTTCACCCGAAATGATTCAGAAAAAGACAG--CTTGTATGACATTGATTTGAA
1211	CATCAGCGTTCGACCTTTCAGACCCGAAAGAGACGCGTGTCTATGACATTGACCTGAAG
1334	GTGGATGACACCTTGAA
1271	GTGGAGAGACCATTTAA
RESULT 8	
ABL02583	
ID	ABL02583 standard; cDNA; 2010 BP.
AC	ABL02583;
DT	26-MAR-2002 (first entry)
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 2231.
KM	Drosophila; developmental biology; cell signalling; insecticide;
OS	pharmaceutical; gene; ss.
PN	Drosophila melanogaster.
MO	MO200171042-A2.
PD	27-SEP-2001.
PF	23-MAR-2001; 2001WO-US09231.
PR	23-MAR-2000; 2000US-191637P.
XX	11-JUL-2000; 2000US-0614150.
XX	(PEKE ) PE CORP NY.
XX	Venter JC, Adams M, Li PWD, Myers EW;
XX	WPI; 2001-656860/75.
XX	P-PSDB; ABB58480.
XX	New isolated nucleic acid detection reagent for detecting 1000 or more
XX	genes from Drosophila and for elucidating cell signalling and cell-cell
XX	interactions -
XX	Claim 1; SEQ ID NO 2231; 21bp + Sequence Listing; English.
XX	The invention relates to an isolated nucleic acid detection reagent
XX	capable of detecting 1000 or more genes from Drosophila. The invention is
XX	useful in developmental biology and in elucidating cell signalling and
XX	cell-cell interactions in higher eukaryotes for the development of
XX	insecticides, therapeutics and pharmaceutical drugs. The invention
XX	discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX	sequences (AB101840-AB16175) and the encoded proteins
XX	(AB57737-AB812072).
XX	The sequence data for this patent did not form part of the printed
XX	specification, but was obtained in electronic format directly from WIPO
XX	at ftp.wipo.int/pub/published_pct_sequences.
XX	Sequence 2010 BP; 509 A; 550 C; 545 G; 406 T; 0 other;
Query Match	14.6%; Score 414.2; DB 23; Length 2010;
Best Local Similarity	68.5%; Pred. No. 1.5e-103;
Matches 620; Conservative	0; Mismatches 273; Indels 12; Gaps 3;
517	AAAGAAAAGAAAGAGCTGCTGACAAATTTCTACTTAAGAGATTGCTGAACCTGATACAGCA
580	AAAGAAAAGAAAGAGCTGCTGACAAATTTCTACTGCGCAGAAAGGTCGGAGATTCGGTCCGCA

OY	577	ATCCAGAGCCATATGATCTCTTGAGCTTTTGAAGAAGAACTGGACAGACTATCATGAG	636
Db	640	ATCCAGAGCCATATGATCTCTTGAGCTTTGAGCGAAAACTGGATGCCATTAATGCC	699
OY	637	GAAGCGCTAGATATCCAAAGAGCGCTTGAACGTCCTCATTAAGCCAAAAAGAAAGCTGCC	696
Db	700	CAAAAGCTGGACATCCAGGAGGCGCTCAAGCGGCCCATGAGAGAGAAAGCGCAAGCTGCC	759
OY	697	AATTTTCATTTCTTAACCTTTTCAATCCGGCTAAGTCAGATGCCGAGATGG---GGAAAG	753
Db	760	CATTTTCATCTCGAACACATTTCTATCCAGCAAGAGGCCACAAATGATGGCCAGGAGGG	819
OY	754	GAGGTGAGCTTCCTGGGAGCGTGGGTGAGAAAGAGCGGCTCTGAGAGATTGACGCTTGTCT	813
Db	820	CGCGCTTGCTCTTTGGAAATTGGCGGGTGGAGGGTCGCTCTTTAGAGGAT-----GGCAA	873
OY	814	CAAAATGATGATGCCACTTAACAAAAAGAGAAAGTTCTCTTCTCTTTAAAGTCCTGTGAT	873
Db	874	GGGGGATGCCAATATCAAAAGATCAAAAGCCAGTTTGTGCTTTTCAAGTCGCTGGTTAT	933
OY	874	TGAACTGAGCAAAAGACCTGTATGGGCGCAGAACCATCTGTGTAATGCGACAGAGCGC	933
Db	934	CGAGCTGATATAGAACTGTACGGTCCGGACAACATCTGTCGAGTGGCATCGCACTCA	993
OY	934	CACATACCCAGAGAGCCGATGGCTTCCAGGTGAAGCGGCCAGAGATCTGATATACGGTG	993
Db	994	CACCAACCCAGAGAGCGAGCGGTTCCAGGTGAAGCGGCCGGGATCGCAATGTGGCTG	1053
OY	994	TACTGTCTTACTGATGTGTGATTACAGCCTCCCGATTAAATTAGACCCCGCGCTAGC	1053
Db	1054	CACCATCTCTCTGTGCTTTGACTACAGCGCGCTGCACTTCAAGCTGGAACCCGCGACTGCC	1113
OY	1054	TCGACTCCTGGGCATTCATACCCAGACTGTCGATGATCATCCAGCACTGTGGCAATA	1113
Db	1114	CAGGTGCTGGCGGTACACACAGACCAAGCGGTCGATCATATCCGCTGTGGCAGTA	1173
OY	1114	TATTTAAGACATATAGCTCCAGAGACCTCACAGAGCGGGAATTGTCACTGTGACAAGTA	1173
Db	1174	CATTAAGACCAACAGCTTACAGAGATGCCACAGCGGAGATACATCAATTGGCACAAGTA	1233
OY	1174	CCTGCAGCAGATCTTTGAGACTCAACGATATGAAGTTTTCAGAGATCCCTCAGCGGCTCCA	1233
Db	1234	CTTGAGACAGATATCTCAGTTGCCAGCGGATGAAAGTTTCCGAGATACCGAGCGCTCAA	1293
OY	1234	TGCCTTGCTTATGCGACACAGAACCTTATCATATTAATCATGTCACAGTGTTCGACCCGAA	1293
Db	1294	TCCGCTCTCGATCCGCGCGATCCGATGTGATCAATCATTTTCAT---TAGAGCGGTGC	1350
OY	1294	TGATCAGAAAAAGACAGCTTGTATATGACATTTGATGTTGAAGTGAGTACACCTTGAAGAC	1353
Db	1351	AGAGAAACAGACAGACTCCCTGTATACGACATCGATGTGAGAGTGGATGATACGCTCAAAA	1410
OY	1354	CCAGATGAATTTCTTTTCTGTCTGCATCTGCACAGCAACAGAGATTGCTACTAGACAA	1413
Db	1411	CCAGATGAACAGCTCTTGTATGAGCACTGCGAGCCAAACAGAGATCCAGGGGCTGCACAC	1470
OY	1414	CAAGA 1418	
Db	1471	CAAGA 1475	
RESULT 9			
ABL02582			
ID	ABL02582	standard; cdna; 4138 BP.	
AC	ABL02582;		
XX	26-MAR-2002	(first entry)	
XX	Drosophila melanogaster	expressed polynucleotide seq ID NO 2228.	
XX	Drosophila;	developmental biology; cell signalling; insecticide;	
XX	pharmaceutical; gene; ss.		
KW			





Claim 1: SEO ID No 357; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP4328) and to cDNAs encoding them (ABQ54131-ABQ56305), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis), systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents cDNA encoding a human ovarian antigen of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published/pct/sequences](http://ftp.wipo.int/pub/published/pct/sequences).

Sequence 1970 BP: 463 A; 590 C; 522 G; 389 T; 6 other; 11 missing

Query Match	14.0%;	Score 397;	DB 24;	Length 1970;
Best Local Similarity	69.3%;	Prod. No. 8.5e-99;		
Matches 615;	Conservative 3;	Mismatches 223;	Indels 46;	Gaps 4
Qy	323	CCGAGAC	CAGGTATGTTGCCAGG	CAGCCGAATGACACCTCAGGACCTTCATGGGACCC 382
Db	172	CAGGCC	CCCGGATGCCGTC	TGGAGCCGGATGCCACAGGGGGCCCATGGGCC 231
Qy	383	CTTGGC	-----TATGGGGGAA	CCCTTCAGTCCGACCTGGCTGGCCCGCAGTCAGGGATG 436
Db	232	CCGGCT	CCTCCCGTACATGGG	CAGCCCGCCGCTGGACCCGCTGGCCCGCGGGCATG 291
Qy	437	GATCAG	TCCCGCAAGAGACCTG	CCCCCTCAG-----CAGATCCAGCAG 478
Db	292	GAGCCG	CCCGCAAGCAGCAGCG	CCCCCGCGGAGAGCGAGGCACAGCCAGGCC 351
Qy	479	GTCAG	CAGCAGGGCGGTCCAA	AAATCGAAACCAACCAATGAAAGAAAGATGGCTGAC 538
Db	352	CAGCGG	TGCCACCGCCCGCGG	AGCCGCGAGTGCACAGAGAGGAGGAGATGGCTGAC 411
Qy	539	AAATTC	TACTCAAGAGATTCTG	TGAACGTGTACAGAAATCCAGGCCCTATATGATCTC 598
Db	412	AAATTC	CTCCTCAAGAGATT	CGGAGCTGTGCCCGAGTCCCAGGCTTATATGACCTC 471
Qy	599	TTGGCT	TTTGAAGGAACTGGAC	CCAGACTATCATAGGAAACCGCTAGATATCAAGAG 658
Db	472	TTGCA	TTTGAGAGAACTGGAT	CAACCATCATGCGGAACGGGTGCACATCCAGGAG 531
Qy	659	GCCTTG	AAACCGTCCCATTAAG	CAAAAACGGGAAGCTGCGAAATTTTCATTCTTAACACTTTC 718
Db	532	GCTCTG	AAGAGGCCCATGAAG	CAAAAACGGGAAGCTGCGACTCTATATCTCCAACACTTTT 591
Qy	719	AATCCG	CTAGTCAGATGCCG	AGATGGGAAGCGGCTGGCTTCCTGGGAGCTTCGG 778
Db	592	AACCTC	GGAAG-CTGATGCT	GAGGATTCGACGCGCAGCATTCCTCTCTGGGAGCTACGG 650
Qy	779	GTAGA	GAGACGGCTCCTG	GGAGATTACGCTTGTCCAAATATGATGCCACTAAACAAAG 838

CC can be used for developing agents for the diagnosis and treatment of  
CC disorders involving unwanted cell proliferation, such as neoplasia,  
CC dysplasia or hyperplasia.

XX Sequence 452 BP; 99 A; 124 C; 104 G; 122 T; 3 other;

Query Match 13.7%; Score 388.8; DB 21; Length 452;  
Best Local Similarity 97.9%; Pred. No. 7.2e-97;  
Matches 415; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 566 CTGTGACCAAGATCCAGGCTTATATGATCTCTGCTTTGAAAGAACTGAGCCAG 625  
DB 423 CAGGTACCAAGATCCAGGCTTATATGATCTCTGCTTTGAAAG-ACCTGAGCCAG 365  
QY 626 ACTATCATGAGAAACGGCTAGATATCCAGAGGCTTGAACGCTCCCTTAAGCAAAA 685  
DB 364 ACTATCATGAG-ACGGCTAGATATCCAGAGGCTTGAACGCTCCCTTAAGCAAAA 306  
QY 686 CCGAAGCTGCGAATTTCTATTTCTAACACTTTCACGCTAAGTCAGATCCGAGAT 745  
DB 305 CCGAAGCTGCGAATTTCTATTTCTAACACTTTCACGCTAAGTCAGATCCGAGAT 246  
QY 746 GGGGAAGGAGCGGTGCTTCTGAGAGCTTGGGTAGAAAGACGGCTCTGAGAGATTCA 805  
DB 245 GGGGAAGGAGCGGTGCTTCTGAGAGCTTGGGTAGAAAGACGGCTCTGAGAGATTCA 186  
QY 806 GCTTGTCCAAATATGATGCCACTAAACAAAGAGAAATCTCTCTCTTTTAAGTTC 865  
DB 185 GCTTGTCCAAATATGATGCCACTAAACAAAGAGAAATCTCTCTCTTTTAAGTTC 126  
QY 866 TTGATGATGAACTGACAAAGACCTGTATGGGCCAGAACACATCTGTAGATGGCAC 925  
DB 125 TTGATGATGAACTGACAAAGACCTGTATGGGCCAGAACACATCTGTAGATGGCAC 66  
QY 926 AGACCGCCACTACCCAGAGAGCCGATGCTTCCAGGTAGAGCGGCCAGAGATGTGAT 985  
DB 65 AGACCGCCACTACCCAGAGAGCCGATGCTTCCAGGTAGAGCGGCCAGAGATGTGAT 6  
QY 986 GTAC 989  
DB 5 GTAC 2

RESULT 13  
AAZ80561/c  
ID AAZ80561 standard; cDNA; 690 BP.

AC AAZ80561;  
XX  
AC  
XX  
DT 07-APR-2000 (first entry)

DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:645.  
XX  
XX Human; gene expression product; diagnosis; tumour; colon cancer;  
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
KW hyperplasia; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO9964576-A2.  
PN  
XX  
XX 16-DEC-1999.  
PD  
XX  
XX 09-JUN-1999; 99WO-1B01062.  
PF  
XX  
XX 10-JUN-1998; 98US-0088801.  
PR  
XX  
XX (FARB ) BAYER CORP.  
PA  
XX  
XX Endege WO, Steimann KE, Astle JH, Burgess CC, Bushnell SE;  
PI Carroll E, Catino TV, Derci A, Ford DM, Lewis ME, Monahan JB;  
PI Schlegel R;

XX WPI; 2000-087220/07.  
DR

PT Novel nucleic acids, used to develop products for the diagnosis and  
PT treatment of disorders involving unwanted cell proliferation,  
PT particularly cancers, especially colon cancer

PS Claim 15; Page 387; 469pp; English.

CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from  
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
CC cDNA clones can be used to generate antisense oligonucleotides which  
CC can be used for antisense therapy. Methods and products from the present  
CC invention can be used for identifying and/or classifying cancerous cells  
CC present in a human tumour, particularly in solid tumours, e.g.  
CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones  
CC can be used for developing agents for the diagnosis and treatment of  
CC disorders involving unwanted cell proliferation, such as neoplasia,  
CC dysplasia or hyperplasia.

SQ Sequence 690 BP; 165 A; 168 C; 170 G; 169 T; 18 other;

Query Match 10.9%; Score 310.2; DB 21; Length 690;  
Best Local Similarity 99.0%; Pred. No. 4.7e-75;  
Matches 312; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1409 GACACCAAGACATGATGATGTGTGGTAAACCAGAGAGAGCGCCGAGCTGAGTTC 1468  
DB 315 GACCTCAAGACATGATGATGTGTGGTAAACCAGAGAGAGCGCCGAGCTGAGTTC 256  
QY 1469 TACTTCCAGCCCTGGGCTCAGAGAGGCTGTGACCATCTTCTACTCCAGGTGAGCAG 1528  
DB 255 TACTTCCAGCCCTGGGCTCAGAGAGGCTGTGACCATCTTCTACTCCAGGTGAGCAG 196  
QY 1529 AGACGACAAAGATTAGACCAAGCCCTGGAATCCGAAATACATAGAGGCTCTCCACAGC 1588  
DB 195 AGACGACAAAGATTAGACCAAGCCCTGGAATCCGAAATACATAGAGGCTCTCCACAGC 136  
QY 1589 CCTGATTCGACGACCAATTTCTTATTTGGGCTGTGCTGCTCTCATATGATCTG 1648  
DB 135 CCTGATTCGACGACCAATTTCTTATTTGGGCTGTGCTGCTCTCATATGATCTG 76  
QY 1649 CTTGCTCTGGGCTGAGGAGGCTTCCAGGGAGTCTGTGATTCACAGACCAAGATG 1708  
DB 75 CTTGCTCTGGGCTGAGGAGGCTTCCAGGGAGTCTGTGATTCACAGACCAAGATG 16  
QY 1709 AAGAGGCTTCACAA 1723  
DB 15 AAGAGGCTTCACAA 1

RESULT 14  
ABX34564  
ID ABX34564 standard; cDNA; 2581 BP.

AC ABX34564;  
XX  
AC  
XX  
DT 13-FEB-2003 (first entry)

DE Human mddt cDNA SEQ ID 125.  
XX  
XX  
XX MDDT; human; disease detection and treatment molecule polypeptide;  
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
KW haemostatic; nephrotropic; antineoplastic; hepatotropic;  
KW gene therapy; protein replacement therapy; cell proliferative disorder;  
KW cancer; adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;  
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
KW psoriasis; hepatitis; gene; ss.  
XX  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200279449-A2.  
PN



XX 23-MAY-2002; 2002WO-IB02939.  
PF  
XX 23-MAY-2001; 2001US-292526P.  
PR  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PA  
XX Auwerx J, Gelman L, Debril M, Rocchi S, Picard F;  
PI  
XX MPI; 2003-120789/11.  
DR  
XX P-PSDB; AAE34721.

XX New human PPAR or RXR cofactor polypeptide, useful for identifying  
PT modulators of nuclear receptor activity, and for treating or preventing  
PT pathologic conditions associated with cell types that express PPAR  
PT receptors, e.g. cancer

XX Example 1; Column 55-56; 31pp; English.

XX The invention relates to human nuclear receptor cofactor polypeptides.  
CC The invention particularly relates to proliferator-activated receptor  
CC (PPAR)-gamma and retinoid X receptor (RXR) cofactor polypeptides.  
CC Polypeptides of the invention are useful for screening compounds that  
CC modulate the interaction of the nuclear receptor with the nuclear  
CC receptor cofactor. Modulators of the invention are useful in preparing  
CC a pharmaceutical composition for treating and/or preventing diseases  
CC or pathologic conditions associated with cell types that express PPAR  
CC receptors. The pathologic conditions treated include metabolic or cell  
CC proliferative disorders such as diabetes, cardiovascular disorders  
CC (e.g. atherosclerosis), renal diseases, neurodegenerative diseases  
CC (e.g. Parkinson's disease, Alzheimer's disease), inflammatory diseases  
CC (e.g. psoriasis, acne), wounds, osteoporosis, infections (e.g. HIV)  
CC or cancer. The invention is useful in gene therapy and as vaccines.  
CC The present sequence is human BRG1-associated factor (BAF) 60C3 DNA.  
CC This sequence is used in the exemplification of the invention.

SQ Sequence 755 BP; 159 A; 240 C; 247 G; 109 T; 0 other;

Query Match 7.8%; Score 221.6; DB 25; Length 755;

Best Local Similarity 68.7%; Pred. No. 1.5e-50;

Matches 346; Conservative 0; Mismatches 134; Indels 24; Gaps 2;

OY 325 GAGACCAGGTATGTTCCAGGACCGCAATGACACTTCAGGGACCTTCATGGGACCC 384  
DB 78 GGGCCCGGAGTGCCTCTGAGCCGAGTCCCAACAGGGGCGCCCATGGCCCC 137  
OY 385 TGGC-----TATGGGGGAACCTTCAGTCCGACCTGGCCAGTCAAGGATGGA 438  
DB 138 GGGCTCCCGTACATGGGACGCCCGCGTGGCAACCGGCTGGCCCCCGGGCATGGA 197  
OY 439 TCAGTCCGCAAGAGACTGCTCCCTCAG-----CAGATCCAGCAGGT 480  
DB 198 GCCCGCCCGCAAGCGACGCGCCCGCGGCGACAGCCAGGCAAGCCAGGGCCA 257  
OY 481 CCAGCAGCAGGGGGTCCAAATGCAACCAATGCAAGAAAGAAAGATGGCTGACAA 540  
DB 258 GCGGTGCCCAAGCGCCCGCGGAGCGCAGTCCCAAGAGGAGAAATGGCTGACAA 317  
OY 541 AATTCTACTCAAGAGATTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 600  
DB 318 AATTCCTCCTCAAGAGATTGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 377  
OY 601 GGGTTTGAAGAGAACTGAGCAGACTATCATGAGAAAGCGCTAGATATCCAGAGGC 660  
DB 378 GGCATTGAGAGAACTGAGTCAATCATGCGAAAGCGGTGACATCCAGAGAGC 437  
OY 661 CTGGAAGCTCCATTAAGAAAGGAAAGCGAATTTTATTTTAAACATTGCA 720  
DB 438 TCTGAAGAGCCCATTAAGAAAGGAAAGCGAATTTTATTTTAAACATTGCA 497  
OY 721 TCCGCTAAGTCAATGCGAGATGGGAGAGGAGCGTGGCTTCTGGAGGCTTGGGT 780  
DB 498 CCTGGGAAGCGTATGCTGAGGATTCGACGCGCAAGCATTTGCTCTGGGAGCTACGGGT 557

OY 781 AGAAGACGGCTCTGTGAGATTC 804  
DB 558 GGAGGGGAAGCTCTGTGATATCC 581

Search completed: October 11, 2003, 12:41:45  
Job time : 525 secs

**THIS PAGE BLANK (USPTO)**



QY	566	CTGTGTAACGAATCCGAGGCTAATAATGATCTCTTGGCTTTGAAAGAACTGGACCAG	623
	423	CAGGTATCCGAATCCGAGGCTAATAATGATCTCTTGGCTTTGAAAGG - AACTGGACCAG	365
Db			
QY	626	ACTATCATAGAGAAACGGCTAGATATCCAGAGGCTTGAAACGTCCTCATTAACCAAAA	685
Db	364	ACTATCATAGAG - AACGGCTAGATATCCAGAGGCTTGAAACGTCCTCATTAACCAAAA	308
QY	686	CGGAAGCTGGAATTTTCATTCTTAACACTTCAATCCGCGCTAAGTAGATGCCGAGAT	745

Db 305 CGGAAGCTCGCAATTTTCATTTTAACTTTTAAATCCGGCTAAGTCAGATGCCGAGGAT 246  
Qy 746 GGGGAAGGACGGTGGCTTCTCGGAGCTTCGGGTAGAAGACGGCTCCTCGAGGATTCA 805  
Db 245 GGGGAAGGACGGTGGCTTCTCGGAGCTTCGGGTAGAAGACGGCTCCTCGAGGATTCA 186  
Qy 806 GCCTTGTCCAAATATGATGCCACTAAACAAAGAGGAAATGTTCTCTCTCTTTTAAGTCC 865  
Db 185 GCCTTGTCCAAATATGATGCCACTAAACAAAGAGGAAATGTTCTCTCTCTTTTAATCC 126  
Qy 866 TTGTTGATTGAATCGACAAGACCTGTATGGGCCAGACAACCATCTCTGTAGATGGCAC 925  
Db 125 TTGTTGATTGAATCGACAAGACCTGTATGGGCCAGACAACCATCTCTGTAGATGGCAC 66  
Qy 926 AGGACCGGCACCTACCACGAGACCGATGGCTTCCAGGTGAAGCGCCAGGAGATGTGAAT 985  
Db 65 AGGACCGGCACCTACCACGAGACCGATGGCTTCCAGGTGAAGCGCCAGGAGATGTGAAT 6  
Qy 986 GTAC 989  
Db 5 GTAC 2

## RESULT 2

US-09-328-111-645/c

; Sequence 645, Application US/09328111

; Patent No. 6262333

; GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.

; APPLICANT: Steinmann, Kathleen E.

; APPLICANT: Astle, Jon H.

; APPLICANT: Burgess, Christopher C.

; APPLICANT: Bushnell, Steven E.

; APPLICANT: Carroll III, Eddie

; APPLICANT: Catino, Theodore J.

; APPLICANT: Derti, Adnan

; APPLICANT: Ford, Donna M.

; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.

; APPLICANT: Schlegel, Robert

; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION

; FILE OF INVENTION: PRODUCTS

; FILE REFERENCE: CCD-257 (US)

; CURRENT APPLICATION NUMBER: US/09/328,111

; CURRENT FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: US 60/088,801

; EARLIER FILING DATE: 1998-06-10

; NUMBER OF SEQ ID NOS: 850

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 645

; LENGTH: 690

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(690)

; OTHER INFORMATION: n = A,T,C or G

US-09-328-111-645

Query Match 10.9%; Score 310.2; DB 3; Length 690;

Best Local Similarity 99.0%; Pred. No. 9e-78;

Matches 312; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1409 GACACACAGACAAATGACTGATGTGGTGAACCCAGAGGAGGAGCGCGAGCTGAGTTC 1468  
Db 315 GACCTCAAGACAAATGACTGATGTGGTGAACCCAGAGGAGGAGCGCGAGCTGAGTTC 256

Qy 1469 TACTTCCAGCCCTGGCTCAGGAGCTGTGTGCCGATATCTTACTTCCAAAGTGCAGCAG 1528  
Db 255 TACTTCCAGCCCTGGCTCAGGAGCTGTGTGCCGATATCTTACTTCCAAAGTGCAGCAG 196

Qy 1529 AGACGACAAAGAAATTAGACAAAGCCCTGGGAATCCGGAATACATAGGGCTCTTCCACAGC 1588  
|||||

Db 195 AGACGACAAAGAAATTAGAGCAAGCCCTGGGAATCCGGAATACATAGGGCTCTTCCACAGC 136  
Qy 1589 CCTGATTGCGACTGCACCAATTTCTTGATTTGGGCCCTGTGCTGCTGCTCATAGTATCTG 1648  
Db 135 CCTGATTGCGACTGCACCAATTTCTTGATTTGGGCCCTGTGCTGCTCATAGTATCTG 76  
Qy 1649 CCTTGGTCTTGTGTTGGGGCGTTCAGGGGATGCTTGTGTTTCAAGGACAAAGCAGCAATG 1708  
Db 75 CCTTGGTCTTGTGTTGGGGCGTTCAGGGGATGCTTGTGTTTCAAGGACAAAGCAGCAATG 16  
Qy 1709 AAGAGGTCTTCACAA 1723  
Db 15 AAGAGGTCTTCACAA 1

## RESULT 3

US-09-397-787-279

; Sequence 279, Application US/09397787

; Patent No. 6468758

; GENERAL INFORMATION:

; APPLICANT: Benson, Darin R.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: King, Gordon E.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN

; FILE OF INVENTION: CANCER THERAPY AND DIAGNOSIS

; FILE REFERENCE: 210121.466C2

; CURRENT APPLICATION NUMBER: US/09/397,787

; NUMBER OF SEQ ID NOS: 1999-09-16

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 279

; LENGTH: 450

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(450)

; OTHER INFORMATION: n = A,T,C or G

US-09-397-787-279

Query Match 2.7%; Score 76; DB 4; Length 450;

Best Local Similarity 61.8%; Pred. No. 1.3e-11;

Matches 168; Conservative 0; Mismatches 80; Indels 24; Gaps 2;

Qy 323 CCGAGACGAGTATGTTGCCAGGCAGCGGATGACACTCAGGACCTTCCATCGGACCC 382  
Db 178 CAGCGCCCGGGATGCCGTCTGGAGCCGGATGCCACCAGGGGGGCCCATGGGCC 237

Qy 383 CCTGGC-----TATGGGGGAACCTTTCAGTCCGACCTGGCCTGGCCCTCAGTCAAGGATG 436  
Db 238 CCGGGCTCCCGGTACATGGGCAGCCCGCGCTGCGACCCGGCTGGCCCCCGGGCATG 297

Qy 437 GATCAGTCCCGCAAGACACTGCCCTCAG-----CAGATCAGCAG 478  
Db 298 GAGCCCGCCGCAAGCAGCAGCGCCCGCCCGCGGAGCCAGGCACAGCAGCCAGGC 357

Qy 479 GTCCAGCAGCAGGGGTCCAAATCGAAACCAATCGAAATCGAAAGAAAGAGATGGCTGAC 538  
Db 358 CAGCCGCTGCCACCCGCCCGCGGAGCGCAGTGCCTCAAGAGGAGGAGATGGCTGAC 417

Qy 539 AAAATCTTACCTCAAAGGATTCGTGAACCTGTT 570  
Db 418 AAAATCTTCCCTCAAAGGATTCGGGAGCTGGT 449

## RESULT 4

US-09-313-294A-3534

; Sequence 3534, Application US/09113294A

; Patent No. 6476212

; GENERAL INFORMATION:

; APPLICANT: Lalgudi, Raghunath V.

; APPLICANT: Ito, Laura Y.

```

: APPLICANT: Sherman, Bradley K.
: TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
:
: FILE REFERENCE: PL-0017 US
:
: CURRENT APPLICATION NUMBER: US/09/313,294A
:
: CURRENT FILING DATE: 1999-05-14
:
: NUMBER OF SEQ ID NOS: 7600
:
: SOFTWARE: PERL Program
:
: SEQ ID NO. 3534
:
: LENGTH: 283
:
: TYPE: DNA
:
: ORGANISM: Zea mays
:
: FEATURE:
:
: NAME/KEY: misc feature
:
: OTHER INFORMATION: Incyte ID NO. 6476212 700611809H1
US-09-313-294A-3534

```

Query Match	2.4%	Score 69.6;	DB 4;	Length 283;
Best Local Similarity	53.3%;	Pred. No. 6.5e-10;		
Matches 147;	Conservative 0;	Mismatches 129;	Indels 0;	Gaps 0;

Oy	1077	AGACGCGCCAGTGAATCATCTCCAGACGCTGGCAATATATTAAACATAGCTCCAG	1130
Db	5	ACACTCGGGCAGCGGTGATGCTGCGCCCTCGCAGATATATTAAACCAAGAACTTCAGA	64
Oy	1137	ACCCTCAGAGGGAGTTTGTCATCTGTGACAAAGTACCCTGACAGACTCTTTGAGACT	1196
Db	65	ACCCAAATGATCTTCTCTTTCATGATGTGACCTCAATTGAAAGAGTGTTGGGGAG	124
Oy	1197	AACGTAATAAGTTTCAGAGATCCCTCAGCGGCTCCATGCGCTGTATTGCAACAGCAAC	1256
Db	125	ACAACTCAAGTTTGCAATGCTGTACAGAGGATATCTTAGCATTTGCTGCCCCGCAC	184
Oy	1257	CTATCATCATTAATCATGCTCATCACTGTTTGAACCCGAATGATCAGAAAAAGACAGCTTGT	1318
Db	185	CCATCAATTTGGAGCAACAAGATTAAAGTGTCGGGAAATGAAGCCATGACAGTGCTTCT	244
Oy	1317	ATGACATTGATGTTGAAGTGCATGACACTTGGAAGA	1352
Db	245	ATGATGTCGTGATGATGTTCCCTTCCGCTGCAGA	280

RESULT 5  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFFLINGER, F.  
APPLICANT: FALZNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.23  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:

```

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpc-F18
US-08-232-463-14

Query Match      2.4%; Score 66.8; DB 1; Length 7218;
Best Local Similarity 7.8%; Pred. No. 2,3e+08;
Matches 35; Conservative 232; Mismatches 179; Indels 0; Gaps 0

```

Query Match	2.4%;	Score 66.8;	DB 1;	Length 7218;
Best Local Similarity	7.8%;	Pred. No. 2.3e-08;		
Matches 35;	Conservative 232;	Mismatches 179;	Indels 0;	Gaps 0;

[illegible]

RESULT 6  
 US-09-103-840A-1/C  
 : Sequence 1, Application US/09103840A  
 : Patent No. 6284328  
 : GENERAL INFORMATION:  
 : APPLICANT: FLEISCHMAN, Robert D.  
 : APPLICANT: WHITE, Owen R.  
 : APPLICANT: FRASER, Claire M.  
 : APPLICANT: VENTER, John C.  
 : TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 : FILE OF INVENTION: TUBERCULOSIS  
 : PILE REFERENCE: 24366-20007.00  
 : CURRENT APPLICATION NUMBER: US/09/103,840A  
 : CURRENT FILING DATE: 1998-06-24  
 : NUMBER OF SEQ ID NOS: 2  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 1  
 : LENGTH: 4411529  
 : TYPE: DNA





Db 1928 CACGGGCACTGGCACCACCGGACCGGCAAGGSCACCGACCGCGGGGGCGC 1869  
Qy 164 GGGTTTCAGTCTGTGGCTTCAAGCGGCGCGGAGCCTCAGGAGGGGGGGCGGC 223  
Db 1868 GGGCGGAGCGCTGCGCTCTCGGTGAGACCGCTCCGCTTGAGCGGCCACGCGCCTC 1809  
Qy 224 TGCTGCTTGGCGCGCGGGAACCTCGGGGCGCTCTGTGCGAATGGGCGCGC 276  
Db 1808 GAGGCGCAGTGCGCGCGGCGGCGGCGGTCCCGGGGCTTGCGCGGC 1756

RESULT 13  
US-09-816-094-3/c  
; Sequence 3, Application US/09816094  
; Patent No. 6534299  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CLO00536-CIP  
; CURRENT APPLICATION NUMBER: US/09/816,094  
; CURRENT FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 7301  
; TYPE: DNA  
; ORGANISM: Human  
US-09-816-094-3

Query Match 1.7%; Score 49; DB 4; Length 7301;  
Best Local Similarity 50.6%; Pred. No. 0.0025;  
Matches 118; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
Qy 44 GCCGCGCGCGCGCGGGAACAGCGGCGCGCGCGGCGCTCGGGGGCGGAG 103  
Db 4243 GCGGTGCGTCCGCGCGGCGGCGGCGGCGCTAGGCGGCGTCCGGCAGACGACCGG 4184  
Qy 104 TTCGGTTCGGTCTTGTGGCGGTGATCGGCGGCTCCGGGAAGATGGCGCGCGC 163  
Db 4183 CACGGGCACTGGCACCACCGGACCGGCAAGGCGACCGACCGCGGGGGCGC 4124  
Qy 164 GGGTTTCAGTCTGTGGCTTCAAGCGGCGCGCGGCGCTCAGGAGGGCGGCGCGC 223  
Db 4123 GGGCGGAGCGCTGCGCTCTCGGTGAGACCGCTCCGCTTGAGCGGCCACGCGCCTC 4064  
Qy 224 TGCTGCTTGGCGCGCGGGAACCTCGGGGCGCTCTGTGCGAATGGGCGCGC 276  
Db 4063 GAGGCGCAGTGCGCGCGGCGGCGGCGGTCCCGGGGCTTGCGCGGC 4011

RESULT 14  
US-08-125-468-1/c  
; Sequence 1, Application US/08125468  
; Patent No. 5589385  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Michael J.  
; APPLICANT: Lotvin, Jason A.  
; APPLICANT: Strathy, Nancy  
; APPLICANT: Fantini, Susan E.  
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
; TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmid  
; TITLE OF INVENTION: useful therein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07470

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/125,468  
APPLICATION NUMBER: 22-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Tsevdos, Estelle J  
REGISTRATION NUMBER: 31,145  
REFERENCE/DOCKET NUMBER: 31,255-02  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201)831-3241  
TELEFAX: (201)831-3305  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 30001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-125-468-1  
Query Match 1.6%; Score 46.2; DB 1; Length 30001;  
Best Local Similarity 50.2%; Pred. No. 0.034;  
Matches 142; Conservative 0; Mismatches 138; Indels 3; Gaps 1;  
Qy 19 AGTCTCGTGGCTGAGCCTCCCGTGGCGCGCGCGGCGGGAACAGCGGCGCGC 78  
Db 23965 AGCGGGCGGCTGAGCGCGCTGGCCAAAGGCGCGGCGGAGGCGCGGCGGCG 23906  
Qy 79 GGGGCGCTCGGGGGCGGGGAGTTCGGGTTCCGTTCTTCTGCGGCTGATCGCG 138  
Db 23905 TGATGTCGCGCGCGCGTGGACGCTGCTGCCCTGCTGCTGCTGCTGCTGCT 23846  
Qy 139 GCTCGGGAAGATGGCGCGCGCGGCTTCCAGTCTGTGGCTTCCAGCGCGCGCG 198  
Db 23845 GCAGCGCACGCTCAGCCCCCGCGGAGCGGCGGCGGCGGCTTCTGCCCTTCGACGCG 23786  
Qy 199 GAGCTCAGGAGGGG---CGGGCGCGGCTGCTGCTTGGGCGCGGAACTCGGGGCC 255  
Db 23785 ACGCTCTCGGGGCTGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23726  
Qy 256 TCCTGTGGAATGGCGCGCGGCTCCGGGTCAAGGCTGTACCGC 298  
Db 23725 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23683

RESULT 15  
US-08-474-933-1/c  
; Sequence 1, Application US/08474933  
; Patent No. 586410  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, Michael J.  
; APPLICANT: Lotvin, Jason A.  
; APPLICANT: Strathy, Nancy  
; APPLICANT: Fantini, Susan E.  
; TITLE OF INVENTION: Cloning of the biosynthetic pathway for  
; TITLE OF INVENTION: chlortetracycline and tetracycline Formation and cosmid  
; TITLE OF INVENTION: useful therein  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07470  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 14:37:03 ; Search time 525 seconds  
(without alignments)  
14041.636 Million cell updates/sec

Title: U66617  
Perfect score: 2841  
Sequence: 1 GAATTCGCCGCTATCCCATAG.....TGGCACAAGGGGATTC 2841

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1457.4	51.3	2017	14	US-10-102-806-137
2	488.8	17.2	606	10	US-09-920-300A-11
3	488.8	17.2	606	12	US-10-099-926-11
4	488.8	17.2	606	13	US-10-033-528-11
5	388.8	13.7	452	10	US-09-879-536-796
6	310.2	10.9	690	10	US-09-879-536-645
7	263.8	9.3	485	11	US-09-918-995-26465
8	237.8	8.4	474	11	US-09-918-995-26465
9	222.6	7.8	447	11	US-09-918-995-28419
10	174.4	6.1	495	11	US-09-918-995-2832
11	123.4	4.3	310	14	US-10-070-676-232
12	113.6	4.0	285	10	US-09-960-352-12390
13	104	3.7	356	10	US-09-783-590-3701
14	102.6	3.6	420	11	US-09-918-995-34033
15	97.4	3.4	462	11	US-09-918-995-28488
16	91	3.2	440	11	US-09-918-995-27234

17	91	3.2	471	11	US-09-918-995-29605	Sequence 23605, A
18	82.6	2.9	15500	9	US-09-764-869-2109	Sequence 2109, Ap
19	82.6	2.9	15500	14	US-10-091-504-2109	Sequence 2109, Ap
20	79.6	2.8	740	9	US-09-770-149-60	Sequence 60, Appl
21	76	2.7	450	9	US-09-876-889-279	Sequence 279, App
22	75	2.6	8788	10	US-09-764-877-2295	Sequence 2295, Ap
23	75	2.6	8788	10	US-09-860-670-247	Sequence 247, App
24	72.8	2.6	520	9	US-09-864-761-8048	Sequence 8048, Ap
25	60.2	2.1	65	12	US-09-908-975-24426	Sequence 24426, A
26	60	2.1	60	12	US-09-908-975-6184	Sequence 6184, Ap
27	54.2	1.9	500	12	US-09-814-353-12798	Sequence 12798, A
28	54.2	1.9	1559	12	US-09-814-353-19072	Sequence 19072, A
29	52	1.8	956	13	US-10-027-632-31508	Sequence 31508, A
30	51.6	1.8	1437	14	US-10-128-714-2188	Sequence 2188, Ap
31	51.6	1.8	1437	14	US-10-128-714-7188	Sequence 7188, Ap
32	50.6	1.8	2046	14	US-10-156-761-4538	Sequence 4538, Ap
33	50.6	1.8	9025608	14	US-10-156-761-1	Sequence 1, Appl
34	50.4	1.8	152331	13	US-10-095-407-16	Sequence 16, Appl
35	49.8	1.8	154746	12	US-09-827-688-8	Sequence 8, Appl
36	49.4	1.7	10464	9	US-09-957-974-1	Sequence 1, Appl
37	49	1.7	1041	9	US-09-916-790-3	Sequence 3, Appl
38	49	1.7	1275	10	US-09-734-032-1	Sequence 1, Appl
39	49	1.7	1275	13	US-10-016-985-1	Sequence 1, Appl
40	49	1.7	1473	13	US-10-016-985-3	Sequence 3, Appl
41	49	1.7	2598	9	US-09-816-094-1	Sequence 1, Appl
42	49	1.7	2598	14	US-10-233-613-1	Sequence 1, Appl
43	49	1.7	2893	9	US-09-916-790-1	Sequence 1, Appl
44	49	1.7	7301	9	US-09-816-094-3	Sequence 3, Appl
45	49	1.7	7301	10	US-09-734-032-3	Sequence 3, Appl

## ALIGNMENTS

```
RESULT 1
US-10-102-806-137
; Sequence 137, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103P1C1
; CURRENT APPLICATION NUMBER: US/10/102,806
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 137
; LENGTH: 2017
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1295)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-102-806-137
Query Match 51.3%; Score 1457.4; DB 14; Length 2017;
Best Local Similarity 98.2%; Pred. No. 0;
Matches 1548; Conservative 8; Mismatches 10; Indels 10; Gaps 8;
```

```
OY 1272 ATGTCATAGGTTGACCCGATGTCAGAAAAGACGCTTGATGACATGATGTTG 1331
DB 4 ATGTATAGAGTTGATCCCGATGTCAGAAAAGACGCTTGATGACATGATGTTG 63
OY 1332 AAGTGATGACACCTTGAGAACCCAGATGATATCTTCTGCTGTCACCGACCAAC 1391
DB 64 AAGTGATGACACCTTGAGAACCCAGATGATATCTTCTGCTGTCACCGACCAAC 123
```

QY 1392 AGGAGATTGCTACTTAGACAAACAGCAATGACTGATGTGTGGTAAACCCAGAGGAGG 1451  
DB 124 AGGAGATTGCTACTTAGACAAACAGCAATGACTGATGTGTGGTAA-CCARARGARG 182  
QY 1452 AGCGCCGAGCTGAGTTCTACTTCCAGCCCTGGGCTCAGAGGCTGTGTCCGATCTTCT 1511  
DB 183 AGCGCCGAGCTGAGTTCTACTTCCAGCCCTGGG-KCAGGAGGCTGTGTGCCRATCTTCT 241  
QY 1512 ACTCAAGGTGCAGCAGAGACAGCAAGAAATAGACAAAGCCCTGGGAATCCGGAAATACAT 1571  
DB 242 ACTCAAGGTGCAGCAGAGACAGCAAGAAATAGACAAAGCCCTGGGAATCCGGNAATCAT 301  
QY 1572 AGGGCCTCTCCACACGCCCTGATTTCGACTGCACCAATTTCTTGATTGGGCCCTGTGTGC 1631  
DB 302 AGGGCCTCTCCACACGCCCTGATTTCGACTGCACCAATTTCTTGAWTTGGGCCCTGTGTGC 361  
QY 1632 CTGCTCATAGTATCTGCTTTGGTCTTGGTGGGGCGTTCCAGGGATGCTGTGGTTCA 1691  
DB 362 CTGCTCATAGTATCTGCTTTGGTCTTGGTGGGGCGTTCCAGGGATGCTGTGGTTCA 421  
QY 1692 AGGCAAGACAGAAATGAAGGGTCTCACAAGACACCTGTTATCTCTTTTCAACCT 1751  
DB 422 AGGCAACACCAAGAAATGAAGGGTCTCACAAGACACCTGTTATCTCTTTTCAACCT 481  
QY 1752 ATCTCTTCCACCCCGAGCTTCCCTTTGCCCCACAAAGTTCCCATGTGCTGTACCCCTCC 1811  
DB 482 ATCTCTTCCACCCCGAGCTTCCCTTTGCCCCACAAAGTTCCCATGTGCTGTACCCCTCC 541  
QY 1812 CTGTGCTACATAGAGACCTTAGATAGTGTGTAGAGAGAACATGTAGTGTATGAGTG 1871  
DB 542 CTGTGCTACATAGAGACCTTAGATAGTGTGTAGAGAGAACATGTAGTGTATGAGTG 601  
QY 1872 CTTGGAATGGATT-GSCCTCAGGCCAGGTGFTTCAAGGGGACAGCAATCACTGATCTTA 1930  
DB 602 CTTGGAATGGATTGGGCTCAGGCCAGGTGFTTCAAGGGGACAGCAATCACTGATCTTG 661  
QY 1931 CCTTTCCAGAGACCCAGGAGTTGGG---TTTTCGCTCTCTTCCAAAGACTCAGGCCCTGTGGG 1987  
DB 662 CCTTTCCAGAGACCCAGGAGTTGGGAGCTTTTCGCTCTCTTCCAAAGACTCAGGCCCTGTGGG 721  
QY 1988 CACTCTATAAGCTAGTTGATCTTGGCTCTCTGTATAACAGAAATCAATTTCTCTCTCC 2047  
DB 722 CACTCTATAAGCTAGTTGATCTTGGCTCTCTGTATAACAGAAATCAATTTCTCTCTCC 781  
QY 2048 CTCACAGTTTGGAAACAACTCTCCCTTCACTGCTGTCCTGTAGCACTACAGAAACC 2107  
DB 782 CTCACAGTTTGGAAACAACTCTCCCTTCACTGCTGTCCTGTAGCACTACAGAAACC 841  
QY 2108 TGGTTCTT-GGCTCCACTGAGCCCGAGGTCACTCCCCAG-CCTCTGGGTGGCCCTGTGT 2165  
DB 842 TGGTTCTTGGGCTCCACTGGCCCCAGGGTCACTCCCGAGCCCTCTGGGTGRCCTGTGT 901  
QY 2166 CAGTCTCTCTCACTCTTAGTTGGGTCCACATCAGTATTTGGAGTTTGTCTTTATT 2225  
DB 902 CAGTCTCTCTCACTCTTAGTTGGGTCCACATCAGTATTTGGAGTTTGTCTTTATT 961  
QY 2226 GCTCCCTCCAGACACTCCCTGTGCTGCCCTTGTGATTCCTCAGATCTGCCCTAATC 2285  
DB 962 GCTCCCTCCAGACACTCCCTGTGCTGCCCTTGTGATTCCTCAGATCTGCCCTAATC 1021  
QY 2286 CCGGCAATTTGGGTGGGGAATCTTTGCTCTTTCCCTTTTCAGAGCCCGAGGATCTCATCTG 2345  
DB 1022 CCGGCAATTTGGGTGGGGAATCTTTGCTCTTTCCCTTTTCAGAGCCCGAGGATCTCATCTG 1081  
QY 2346 GGGAACTGTCAATGTCAGAGAGGCTGTCTTCTTCGTGAGTTTGGAGATGTGACTCATTC 2405  
DB 1082 GGGAACTGTCAATGTCAGAGAGGCTGTCTTCTTCGTGAGTTTGGAGATGTGACTCATTC- 1140  
QY 2406 CATTCACACTCCACCCCTGCTCTGCAATCCCTTAATGGAGAAACGGGCTTAAACCAAA 2465  
DB 1141 CATTCACACTCCACCCCTGCTCTGCAATCCCTTAATGGAGAAACGGGCTTAAACCAAA 1200

QY 2466 CGGTAATAAAGCCCTGGGCAATCCCTGTCTTCTGCTGCTTGTCTGCCAGTTGACACC 2525  
DB 1201 CGGTAATAAAGCCCTGGGCAATCCCTGTCTTCTGCTGCTTGTCTGCCAGTTGACACC 1259  
QY 2526 TACTGGTGACTTCTAGGGCACTAGAGGAGTGAAGCGCTAGGGCTGGAGAAATAGCGCTGA 2585  
DB 1260 TACTGGTGACTTCTAGGGCACTAGAGGAGTGAAGCGCTAGGGCTGGAGAAATAGCGCTGA 1319  
QY 2586 GTTGGGTGTTGACTCTTCCCTCTCCCTGCTCAGAGATTTGTGACTCCCCCAGGCCCTGTC 2645  
DB 1320 GTTGGGTGTTGACTCTTCCCTCTCCCTGCTCAGAGATTTGTGACTCCCCCAGGCCCTGTC 1379  
QY 2646 CCTCAAAGCTTTCAGACCCCTCAGGTAGCAGAGACCTTGTGATCTTGGCCCCCTTGGATC 2705  
DB 1380 CCTCAAAGCTTTCAGACCCCTCAGGTAGCAGAGACCTTGTGATCTTGGCCCCCTTGGATC 1439  
QY 2706 TGAGATGGTTTTTCATCTTCCAGGAGACCTCACAATCTTCTTCCAGGTTGTATCACC 2765  
DB 1440 TGAGATGGTTTTTCATCTTCCAGGAGACCTCACAATCTTCTTCCAGGTTGTATCACC 1499  
QY 2766 CCCAGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAAGGTTGGAGACAGCTGGG 2825  
DB 1500 CCCAGTTAGCATATCCAGGCTCGCAGACTCAACACAGCAAGGTTGGAGACAGCTGGG 1559  
QY 2826 CACAAAGGGGATTC 2841  
DB 1560 CACAAAGGGGAAATTC 1575

RESULT 2

US-09-920-300A-11/c  
; Sequence 11, Application US/09920300A  
; Patent No. US2020136728A1  
; GENERAL INFORMATION:  
; APPLICANT: King, Gordon E.  
; APPLICANT: Mesgher, Madeleine Joy  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Secrist, Heather  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE OF INVENTION: AND DIAGNOSIS OF COLON CANCER  
; FILE REFERENCE: 210121.547  
; CURRENT APPLICATION NUMBER: US/09/920.300A  
; CURRENT FILING DATE: 2001-07-31  
; NUMBER OF SEQ ID NOS: 1789  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 606  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 365..433, 507, 549, 591  
; OTHER INFORMATION: n = A,T,C or G  
US-09-920-300A-11

Query Match 17.2%; Score 488.8; DB 10; Length 606;  
Best Local Similarity 93.6%; Pred. No. 8.7e-144;  
Matches 553; Conservative 0; Mismatches 30; Indels 8; Gaps 4;

QY 2237 GACACTCCTGTGGCTGCC---TTTGTGATTCCTCAGATCTGCC---TAATCCCGGG 2290  
DB 590 GACACTCCTGTGGCTGCCCTTTGTGAATTCCTTCAGATNTGCCCTAATCCCGGG 531  
QY 2291 CATTTGGGTGGGGAATCTTGCCTTTCCCTTCAGAGCCCGAGGATCTCATCTGGGAA 2350  
DB 530 CATTTGGGGGGGAAATCTTGCCTTTCCCTTCAGAGCCCGAGGATCTCATCTGGGAA 471  
QY 2351 CTGTCAATGGCCAGAGGCTGTTCTTCTCTGCAGTT-TGGAGATGTGACTCATTCATT 2409  
DB 470 CTGTCAATGGCCAGAGGCTGTTCTTCTCTGCTGTTTGGAGATGTGACTCATTCATT 411  
QY 2410 CACTCACTCCACCTGCTCTGCAATCCCTTAATGGAGAAACGGGCTTAAACCAACGGG 2469



Db 350 T-AAAAAGCCCTGGGCAATCCCTGTCTTCCCTTGTCTGCGCAGTTGACACCTACT 292  
QY 2530 GGTGACTTCTAGGCACTGAGAGTGAAGCGCTTAGGCTGGGAATAGCGCTGAGTTG 2589  
Db 291 GGTGACTTCTAGGCACTGAGAGTGAAGCGCTTAGGCTGAAGATAGCGCTGAGTTG 232  
QY 2590 GGTGTTGTGACTTCTCCCTCTCCCTGCTCACAGGATTTGACTCCCCCAGCCCTGCCCTC 2649  
Db 231 GGTGTTGTGACTTCTCCCTCTCCCTGCTCACAGGATTTGACTCCCCCAGCCCTGCCCTC 172  
QY 2650 AAAGCTTCAGACCCCTCAGGTAGCAGAGGACCTTTGTGATCTTGCCCTTTGGATCTCAG 2709  
Db 171 AAAGCTTCAGACCCCTCAGGTAGCAGAGGACCTTTGTGATCTTGCCCTTTGGATCTCAG 112  
QY 2710 ATGGTTTGTGACTTCTTCAGAGAGCCCTCAGATTTCTTCTCCAGGTTGTATCACCCCG 2769  
Db 111 ATGGTTTGTGACTTCTTCAGAGAGCCCTCAGATTTCTTCTTCAGGTTGTATCACCCCG 52  
QY 2770 AGTTAGCATATCCAGGCTCGCAGACTCAACAGCAAGGGTGGGAGACAG 2820  
Db 51 AGTTAGCATATCCAGGCTCGCAGACTCAACAGCAAGGGTGGGAGACAG 1

## RESULT 5

US-09-879-536-796/c  
; Sequence 796, Application US/09879536  
; Patent No. US20020144298A1

## GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie.  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/879,536  
; CURRENT FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/088,801  
; PRIOR FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 796  
; LENGTH: 452  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(452)  
; OTHER INFORMATION: n = A,T,C or G

US-09-879-536-796

Query Match 13.7%; Score 388.8; DB 10; Length 452;  
Best Local Similarity 97.9%; Pred. No. 3.6e-112;  
Matches 415; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 566 CTGTACAGAAATCCAGGCTATATGATCTCTTGGCTTTTGAAGAAACTGACACAG 625  
Db 423 CAGGTACAGAAATCCAGGCTATATGATCTCTTGGCTTTTGAAGG-AACTGACACAG 365  
QY 626 ACTATCATGAGAAACGGCTAGATATCAAGAGGCTTTGAACGTCCTCAATTAAGCAAAA 685  
Db 364 ACTATCATGAGG-AAACGGCTAGATATCAAGAGGCTTTGAACGTCCTCAAGCAAAA 306  
QY 686 CGGAAGCTGCGAAATTTTCATTTTCAATCCGGCTTAAGTCAGATGCCGAGGAT 745

Db 305 CGGAAGCTGCGAAATTTTCATTTTCAATCCGGCTTAAGTCAGATGCCGAGGAT 246  
QY 746 GGGGAAGGGACGGTGGCTTCTCGGAGCTTCGGGTAGAGGACGGCTCCTCGAGGATTC 805  
Db 245 GGGGAAGGGACGGTGGCTTCTCGGAGCTTCGGGTAGAGGACGGCTCCTCGAGGATTC 186  
QY 806 GCCTTGTCCAAATATGATGCCACTTAAACAAAAGAGGAAGTTCTCTTCTTTAAAGTCC 865  
Db 185 GCCTTGTCCAAATATGATGCCACTTAAACAAAAGAGGAAGTTCTCTTCTTTAAATCC 126  
QY 866 TTGTGATTTGAACCTGGGCAAGACCTGTATGGGCCAGACAACCATCTGTGTAGATGGCAC 925  
Db 125 TTGTGATTTGAACCTGGGCAAGACCTGTATGGGCCAGACAACCATCTGTGTAGATGGCAC 66  
QY 926 AGGACCCCACTCCAGGAGACCGATGGCTTCAGGTGAAGCGGCCAGGAGTGTGAAT 985  
Db 65 AGGACCCCACTCCAGGAGACCGATGGCTTCAGGTGAAGCGGCCAGGAGCGTGAAT 6  
QY 986 GTAC 989  
Db 5 GTAC 2

## RESULT 6

US-09-879-536-645/c  
; Sequence 645, Application US/09879536  
; Patent No. US20020144298A1

## GENERAL INFORMATION:

; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie.  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/879,536  
; CURRENT FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/088,801  
; PRIOR FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 645  
; LENGTH: 690  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(690)  
; OTHER INFORMATION: n = A,T,C or G

US-09-879-536-645

Query Match 10.9%; Score 310.2; DB 10; Length 690;  
Best Local Similarity 99.0%; Pred. No. 3.8e-87;  
Matches 312; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1409 GACAACAAGACATGACTGATGTGGTGAACCCAGAGGAGGAGCGCGAGCTGAGTTC 1468  
Db 315 GACCTCAAGACAAATGACTGATGTGGTGAACCCAGAGGAGGAGCGCGAGCTGAGTTC 256  
QY 1469 TACTTCCAGCCCTGGGCTCAGGAGGCTGTGTGGCGATCTTCTACTCAAGGTGCAGCAG 1528  
Db 255 TACTTCCAGCCCTGGGCTCAGGAGGCTGTGTGGCGATCTTCTACTCAAGGTGCAGCAG 196  
QY 1529 AGACACAAGAAATTAGAGCAAGCCCTGGGAATCCGGAATACATAGGGCCCTCTCCACAGC 1588

```

Db      195 AGACGACAAAGATTAGAGCAAGCCCTGGGAATCCGGAATCATAGAGCCCTCCACAGC 136
      1589 CCTGATTGACATGACACCAATTTCTTGAATTTGGGCGCTGCTGCTCCCTATAGTATCTG 1648
      135 CCTGATTGACATGACCAATTTCTTGAATTTGGGCGCTGCTGCTCCCTATAGTATCTG 76
      1649 CCTGATTGCTGCTGGGCGCTTCCAGGGAGTGTGTTCAAGACAAAGACAGATG 1708
      75 CCTGATTGCTGCTGGGCGCTTCCAGGGAGTGTGTTCAAGACAAAGACAGATG 16
      1709 AAGAGGCTCTCAAA 1723
      15 AAGAGGCTCTCAAA 1
      15 AAGAGGCTCTCAAA 1

RESULT 7
US-09-918-995-16164
; Sequence 16164, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16164
; LENGTH: 485
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(485)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-16164

Query Match
Best Local Similarity 73.4%; Pred. No. 1.6e-72;
Matches 337; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

      819 ATGATGCCACTAAACAAAGAGAGTCTCTTCTTTTAAGTCTTGGTGAATGAAC 878
      26 AGGAACCTCAGACAGACAGAGCGGAGTCTCTTCTTTCAAGAGTTGGTCATCGAGC 85
      879 TGGACAAAGACTGTATGGGCCAGACAAACCATTTGTTAGATGGCACAGACCGGACTA 938
      86 TGGACAAAGACTGTATGGGCCAGACAAACCATTTGTTAGATGGCACAGACCGGACTA 145
      939 CCAGAGAGCCGATGCTTCCAGGTGAAGGCGCCAGAGAGTGAATGATACGGTGAAC 998
      146 CCAGAGAGCCGATGCTTCCAGGTGAAGGCGCCAGAGAGTGAATGATACGGTGAAC 205
      206 TCGCTCTTACGTGGAGCTACAGCCCTCCCAAGTTCAAGTGAATCCCGCTTAACCGGC 265
      1059 TCGTGGGACATCCATACCCGAGCTCGCTCCAGTATCATCCAGACATGTGGCATATATTA 1118
      266 TCGTGGGACATCCATACCCGAGCTCGCTCCAGTATCATCCAGACATGTGGCATATATTA 325
      1119 AGACATATAGCTCCAGAGCCCTCAGAGGCGGAGTTGTCAATCTGTGACAAGTACTGC 1178
      326 AGACATATAGCTCCAGAGCCCTCAGAGGCGGAGTTGTCAATCTGTGACAAGTACTGC 385
      1179 AGCAGATCTTTGAGACTCAACGATGATGATTTTCAAGATTCCTCAGCGGCTTCATGCT 1238
      386 AGCAGATCTTTGAGACTCAACGATGATGATTTTCAAGATTCCTCAGCGGCTTCATGCT 445

```

```

      1239 TCGTATGCCACAGAACCTATCATCATTAATCATGTCA 1277
      446 TCGTATGCCACCTGACCCCAATTCATCAACCATGTCA 484

RESULT 8
US-09-918-995-26465
; Sequence 26465, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 26465
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(474)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26465

Query Match
Best Local Similarity 72.5%; Pred. No. 2.7e-64;
Matches 308; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

      899 CCAGAACCACTGTGTAATGAGACAGAGCCGCTAGTCCAGAGACCGATGCTTC 958
      50 CCGAACCACTGTGTAATGAGACAGAGCCGCTAGTCCAGAGACCGATGCTTC 109
      959 CAGGTGAACCGGCGGAGATGTAATGATGAGTACTGCTCTATGCTGATTC 1018
      110 CAGGTGAACCGGCGGAGATGTAATGATGAGTACTGCTCTATGCTGATTC 169
      1019 CAGGTGAACCGGCGGAGATGTAATGATGAGTACTGCTCTATGCTGATTC 1078
      170 CAGGTGAACCGGCGGAGATGTAATGATGAGTACTGCTCTATGCTGATTC 229
      1079 ACTGTCAGATGATCATCCAGACCTGTGCAATATTAAGACATTAAGTCCAGAGC 1138
      230 ACTGTCAGATGATCATCCAGACCTGTGCAATATTAAGACATTAAGTCCAGAGC 289
      1139 CCTCAGAGCGGAGTGTGATCTGTGACAAAGTACTGACAGAGATCTTTGAGACTCA 1198
      290 CCTCAGAGCGGAGTGTGATCTGTGACAAAGTACTGACAGAGATCTTTGAGACTCA 349
      1139 CGTATGAGTTTTCAGAGTCCCTGAGGCGCTCCATGCTTATGACACAGAACT 1258
      350 CGTATGAGTTTTCAGAGTCCCTGAGGCGCTCCATGCTTATGACACAGAACT 409
      1259 ATCATCATTAATCATGATCATAGTGTGACCCGAATGATCAGAAAAGACAGCTTTAT 1318
      410 ATTCATCATTAATCATGATCATAGTGTGACCCGAATGATCAGAAAAGACAGCTTTAT 469
      1319 GACAT 1323
      470 GACAT 474

RESULT 9
US-09-918-995-28419
; Sequence 28419, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.

```

; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 28419  
; LENGTH: 447  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; LOCATION: (1)...(447)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-28419

Query Match 7.8%; Score 222.6; DB 11; Length 447;  
Best Local Similarity 74.1%; Pred. No. 1.7e-59;  
Matches 297; Conservative 0; Mismatches 99; Indels 5; Gaps 1;  
QY 583 GGCCTATATGGATCTCTGGCTTTTGAAGGAAACTGGACCAAGATATCATGAGGAAACG 642  
DB 52 GGAATGCATGGACCTCTTGGCATTTTGAAGGAAACTGGATCAAAACCATCATGCGGAAGCG 111  
QY 643 GCTAGATATCAAGAGGCTTGAACAGTCCCAATTAAGCAAAACGGAAGCTGCGAATTT 702  
DB 112 GGTGACATCCAGGAGGCTCTGAAGAGGCCCAATGAAGCAAAAGCGGAAGCTGCGACTCTA 171  
QY 703 CATTTCTTAACACTTTCAATCCGGCTTAAGTCAGATCCGAGGATGGGAAGGACCGTGGC 762  
DB 172 TATCTCCACACTTTTAAACCTCGCAGACCTGATGCTGAGGATTCGACGCGCAGATTGC 231  
QY 763 TTCTCGGAGCTTCGGGTAGAGGACGGCTCTCGGAGGATTCAGCTTGTCCAAATATGA 822  
DB 232 CTCCTGGGAGTACGGGTGGAGGGAAGCTCTCGGATGAT-----GTACGTCCTCGGCCCA 286  
QY 823 TCCCACTAAACAAGAGGAGTCTCTCTCTTCTTAAGTCTTGGTGGATTAAGTCTGA 882  
DB 287 GCCCAACAACAAGAGGAGTCTCTCTCTTCTTCAAGAGTGTGGTATCGAGCTGA 346  
QY 883 CAAGACCTGTATGGCGCAGACCAACATCTGTAGATGGCAGACGCGCCACCTACCCA 942  
DB 347 CAAAGATCTTATGSCCTTGACCAACACCTCTGTTAGTGGCATCGACACCCACCA 406  
QY 943 GGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAGATGTGA 983  
DB 407 GGAGAGGAGCGCTTCCAGGTGAACAGCGCTGGGACCTGA 447

RESULT 10  
US-09-918-995-2832  
; Sequence 2832, Application US/0918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2832  
; LENGTH: 495  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature

; LOCATION: (1)...(495)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-2832  
Query Match 6.1%; Score 174.4; DB 11; Length 495;  
Best Local Similarity 65.4%; Pred. No. 3.4e-44;  
Matches 291; Conservative 0; Mismatches 142; Indels 12; Gaps 2;  
QY 666 AACGTCCCATTAAGCAAAACGGAAGCTGCGAAATTTTCATTCTTAACACTTTCAATCCGG 725  
DB 50 AAAAGCCTCTGACACAAAAGCGAAGCTTCGGATCTACATTTCCCAATACGTTCAATCCCA 109  
QY 726 CTAAGTCAGATGCCGAGATGGGAAGGACGGTGGCTTCTCGGAGCTTCGGGTAGAAG 785  
DB 110 GCAAGCGGAAGCGGATAGTCAGGAACTGCAGGACCCCTGGGGGAACCCAGCAGGGG 169  
QY 786 ----GACGGCTCTGGAGGATTACGCTTGTCCAAA-----TATGATGCCACTAAAC 833  
DB 170 ACAAGGTGGCTTCTCGGAACCTCCGAGTGGGAAGAAACTGCTGGATGATCCTAGCAAC 229  
QY 834 AAAAGAGGAAGTCTCTCTCTTCTTAAAGTCTTGGTGAATGAACCTGGACAAAGACCTGT 893  
DB 230 AGAAGAGGAAGTCTTCTTCTTCTTAAAGACCTCGTCATTCAGCTGGACGAGGCTGT 289  
QY 894 ATGGGCCAGACCAACCTCTGTAGATGGCAAGGACCGCCACCTACCCAGAGACCGATG 953  
DB 290 ACGGCGCTGACAATCACTCTGGTGGAGTGGCAACCGGATGCCCAACCCAGGAGACAGATG 349  
QY 954 GCTTCCAGGTGAAGCGGCCAGGAGATGATGATGACGGTGTACGGTGTACTGTCTCTATGCTGG 1013  
DB 350 GCTTCCAAGTAAAAACGGCTGGAGACCTCAACGTCGAAGTGCACCTCTCTGCTCATGTGG 409  
QY 1014 ATTACGACCTCCCGAGTTTAAATAGACCCCGCTAGCTCGACTCTCTGGGCTCCATA 1073  
DB 410 ATCATGACCTCCCGAGTACAAATGGACCCCGCTAGTGGCAAGGCTTGTCTGGAGTGACA 469  
QY 1074 CCCAGACTCGTCCAGTGTATCATCCA 1098  
DB 470 CGCAGACAAAGGCGCGNATCATGCA 494

RESULT 11  
US-10-070-676-23/c  
; Sequence 23, Application US/10070676  
; Publication No. US20030059788A1  
; GENERAL INFORMATION:  
; APPLICANT: Toque, Bruno  
; APPLICANT: Bracco, Laurent  
; APPLICANT: Schweighoffer, Fabien  
; TITLE OF INVENTION: Genetic Markers of Toxicity, Preparation  
; TITLE OF INVENTION: and Uses  
; FILE REFERENCE: 50146/003002  
; CURRENT APPLICATION NUMBER: US/10/070,676  
; CURRENT FILING DATE: 2002-03-06  
; PRIOR APPLICATION NUMBER: PCT/FR00/02503  
; PRIOR FILING DATE: 2000-09-12  
; PRIOR APPLICATION NUMBER: FR 99/11405  
; PRIOR FILING DATE: 1999-09-13  
; PRIOR APPLICATION NUMBER: 09/456,370  
; PRIOR FILING DATE: 1999-12-08  
; NUMBER OF SEQ ID NOS: 37  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 23  
; LENGTH: 310  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc feature  
; LOCATION: 3, 33, 34  
; OTHER INFORMATION: n = A,T,C or G  
US-10-070-676-23

Query Match 4.3%; Score 123.4; DB 14; Length 310;

Best Local Similarity 66.0%; Pred. No. 3.7e-28;  
Matches 192; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

QY 1097 CAAGCAGCTGGCAATATATTAGACATTAAGCTCCAGACCCCTGAGGGGGAGTT 1156-  
DB 306 CAGGCTTGGCTTTTACATCAAGCAACAGCTGAGAGGGGACAGGGGAGTAC 247  
QY 1157 GTATCTGTGACAAAGTACCTGAGCAGATCTTTGAGACTCAACGTATGAAGTTTTCAGAG 1216  
DB 246 ATCAATGACACCGTTACTTCCGCCGATC-TTCAGTTGGGCCGATCCGTTTCTCCGAG 188  
QY 1217 ATCCCTCAGCGCTCCATCTTGTATGACACAGAACTATCATTAATCATGTC 1276  
DB 187 ATTCCCATGAAGCTGAGGGGTTGCTGACAGCATCCAGACCCCATTTGTATCAACCATGTC 128  
QY 1277 ATCAGCTGTGACCCGAATATCAGAAAAGACAGCTTTGTTATGACATTGATGTTGAAGT 1336  
DB 127 ATTAGGTGACCTTACACACCAAGAAAGACGCTGTTACGACATGATGTGAGGTG 68  
QY 1337 GATGACACCTTGAAGACCAAGATGAATCTTTCTGCTGCTCCATGCGAGC 1387  
DB 67 GACGACCCACTGAAGGCCCAATGAGCAATTTTNGGCTCTTACCAACATC 17

RESULT 12  
US-09-960-352-12390  
; Sequence 12390. Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Mairen, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 12390  
; LENGTH: 285  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 53-LIB058-001-Q1-K1-F2  
US-09-960-352-12390

Query Match 4.0%; Score 113.6; DB 10; Length 285;  
Best Local Similarity 65.8%; Pred. No. 4.5e-25;  
Matches 192; Conservative 0; Mismatches 79; Indels 21; Gaps 1;

QY 656 GAGGCTTGAAAGCTCCCTTAAGCAAAACGGAAGCTGCGAATTTCTATTCTAACCT 715  
DB 15 GATGCTCTCAAGCTGTTTTCGAAGCAGAAAGCGCAACCTGCTGCTATATCTCCATATCT 74  
QY 716 TTCAATCCGGCTAAGTCAGATGCGGAGATGGGGAAGGGAGCGGTGCTTCTGGGAGCTT 775  
DB 75 TTTAACCTGCTAAGCCGATGCTGAGATTCGAGGGAGCAATGCTCTCTGGGAGCTG 134  
QY 776 CGGGTGAAGAGACGCTCTGAGATTCAGCTTTGTCAAATATGATGCACTAAACAA 835  
DB 135 CTGGTGAAGGGAGGAGCTCTGG-----ATGACCCGAGCTGGCAT 173  
QY 836 AAGAGGAAGTTCTCTTCTCTTTTAAGTCCCTGGTATTTGAATGACAAAGCTGTAT 895  
DB 174 ATGCGGAGATTTCTGCTTCAATCAAGAGCTCGTATGATCTGAGCATAGACCTTTAT 233  
QY 896 GGGCCGAGCAACATCTGTAGATGAGCAAGGACCGGCACTACCAACAGAGA 947  
DB 234 GGGCTGACCAACCTGTGGAGTGGACCTGACACCAACCACTACAGAGA 285

RESULT 13  
US-09-783-590-3701

; Sequence 3701. Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillon, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21  
; NUMBER OF SEQ ID NOS: 12485  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3701  
; LENGTH: 356  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (5)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (8)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (12)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (23)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (44)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (92)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (120)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (138)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (237)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (252)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (268)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (314)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (325)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (327)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (331)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (347)

OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (353)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-3701

Query Match 3.7%; Score 104; DB 10; Length 356;

Best Local Similarity 88.9%; Pred. No. 5.7e-22;

Matches 200; Conservative 0; Mismatches 15; Indels 10; Gaps 8;

Qy 1416 AGACAATGACTGATGTGTGGGTACCCAGAGGAGCGCGGAGCT-GAGTCTTACTTC 1474

Db 22 ANACAATGACTGATGTGTGGNAACCCAAAGGAGGAGCGCGGAGCTGGAGTTTACTTC 81

Qy 1475 CAGCCCTGGCTCAGG-AGGCTGTGCGGATCTTCT-ACTCCAAGGTGAGCAGAGA- 1531

Db 82 CAGCCCTGGNTCAGAGGCTGTGCGGATCTTCTTACTTCCAAAGGTGAGCAGNGAA 141

Qy 1532 -CGACAAGAATTAGAGCAAGCCCTGGGAA--TCCGGAATACATAGGCGCTCT-CCACAG 1587

Db 142 CGACAAGGAATTAGAGCAAGCCCTGGGAAATCCGGTAATACATAGGCGCTCTGCCACAG 201

Qy 1588 CCTGATTCGACT-CCACAATTCTT-GATTTGGCCCTGTGCTG 1630

Db 202 CCTGATTCGACTGGCACCATTCTTGGATTGGNCCTGTGATG 246

RESULT 14

US-09-918-995-34033

Sequence 34033, Application US/09918995

Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995

PRIOR FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 34033

LENGTH: 420

TYPE: DNA

ORGANISM: Homo sapiens

US-09-918-995-34033

Query Match 3.6%; Score 102.6; DB 11; Length 420;

Best Local Similarity 71.4%; Pred. No. 1.8e-21;

Matches 135; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 1009 GCTGATTTACAGCCTCCCGAGTTTAAATTAGACCCCGCTAGCTCAGCTCTGGGCAT 1068

Db 3 GCTGACTACCAAGCCTCCCGAGTTTCAAACTGGATCCCGCTAGCCCGCTGCTGGGCT 62

Qy 1069 CCATACCCAGACTCTCTCAGTGCATCATCAAGCACTGTGGCAATATATTAAGACACATAA 1128

Db 63 GCACACAGAGCCGCTCAGCCATGTTCAGGCCCTGTGGCAGTATGTGAAGACCAACAG 122

Qy 1129 GCTCCAGGACCTTCACGAGCGGGAGTTTGTTCATCTGTGACAGTACCTGACAGCATTT 1188

Db 123 GCTGAGGACTCCCATGACAGGAATACATCAATGGGACAAGTATTTCCAGCAGATTCC 182

Qy 1189 TGAGACTCA 1197

Db 183 TGAGTAGCA 191

RESULT 15

US-09-918-995-28488

Sequence 28488, Application US/09918995

Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.

TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

FILE REFERENCE: 20411-756

CURRENT APPLICATION NUMBER: US/09/918,995

PRIOR FILING DATE: 2001-07-30

PRIOR APPLICATION NUMBER: US/09/235,076

PRIOR FILING DATE: 1999-01-20

NUMBER OF SEQ ID NOS: 38054

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 28488

LENGTH: 462

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)...(462)

OTHER INFORMATION: n = A,T,C or G

US-09-918-995-28488

Query Match 3.4%; Score 97.4; DB 11; Length 462;

Best Local Similarity 73.1%; Pred. No. 8.3e-20;

Matches 125; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 1409 GACAAAGACAATGACTGATGTGGTAAACCCAGAGGAGCGCCGAGCTGAGTTC 1468

Db 280 GACCTCAAGGTGATGACAGATGTAGCCGCAACCTGAAGAGGAGCGCCGCTGAGTTC 339

Qy 1469 TACTTCAGCCCTGGGCTCAGGAGGCTGTGCGGATCTTCTTACTCCAGGTCAGGAG 1528

Db 340 TACCACCAAGCCTGGTCCCGAGGCGGTGCTGCTTCTTACTGCAAGATCCAGCAG 399

Qy 1529 AGACGACAAGAAATTAGAGCAAGCCCTCGGAATCCGGAATACATAGGCGCTC 1579

Db 400 CGCAGGAGGAGCTGGAGCAGTCCCTGCTGTTGCGCAACACCTTAGGAGCCC 450

Search completed: October 11, 2003, 17:35:10

Job time : 532 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 12:21:42 : Search time 3873 Seconds  
(without alignments)  
17828.303 Million cell updates/sec

Title: U66617

Perfect score: 2841  
Sequence: 1 GAAATCCCGCTATCCCATAG.....TGGGCAACAAGGGGATTTC 2841

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_escba:\*  
2: em\_escba:\*  
3: em\_escba:\*  
4: em\_escba:\*  
5: em\_escba:\*  
6: em\_escba:\*  
7: em\_escba:\*  
8: em\_escba:\*  
9: em\_escba:\*  
10: em\_escba:\*  
11: em\_escba:\*  
12: em\_escba:\*  
13: em\_escba:\*  
14: em\_escba:\*  
15: em\_escba:\*  
16: em\_escba:\*  
17: em\_escba:\*  
18: em\_escba:\*  
19: em\_escba:\*  
20: em\_escba:\*  
21: em\_escba:\*  
22: em\_escba:\*  
23: em\_escba:\*  
24: em\_escba:\*  
25: em\_escba:\*  
26: em\_escba:\*  
27: em\_escba:\*  
28: em\_escba:\*  
29: em\_escba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2439.4	85.9	3206	11	BC000063 Homo sapi
2	892	31.4	929	13	BQ33801 AGENCOURT
3	847.8	29.8	933	13	BX386559 BX386559
4	839.4	29.5	1201	13	BX375270 BX375270

5	833.6	29.3	907	13	BQ929257	BQ929257 AGENCOURT
6	825.2	29.0	970	9	AL521992	AL521992 AGENCOURT
7	810.6	28.5	979	13	BQ934447	BQ934447 AGENCOURT
8	809.2	28.5	953	13	BQ845642	BQ845642 AGENCOURT
9	795.6	28.0	2476	11	AK075611	AK075611 Mus muscu
10	779.8	27.4	972	13	BQ881274	BQ881274 AGENCOURT
11	772.4	27.2	872	13	BQ931457	BQ931457 AGENCOURT
12	771.8	27.2	1003	12	BM457092	BM457092 AGENCOURT
13	765	26.9	1201	13	BX363764	BX363764 BX363764
14	763.6	26.9	858	13	BQ557139	BQ557139 AGENCOURT
15	760.2	26.8	878	13	BQ949766	BQ949766 AGENCOURT
16	755.2	26.6	820	9	AU132848	AU132848 AGENCOURT
17	755.2	26.6	1123	12	BQ061434	BQ061434 AGENCOURT
18	748.6	26.3	953	13	BQ716211	BQ716211 AGENCOURT
19	747.6	26.3	825	13	BQ521876	BQ521876 AGENCOURT
20	746.6	26.3	783	9	AU132128	AU132128 AGENCOURT
21	742	26.1	1094	13	BX340735	BX340735 BX340735
22	734	25.8	913	13	BQ591331	BQ591331 AGENCOURT
23	725.6	25.5	882	13	BQ232218	BQ232218 AGENCOURT
24	720.6	25.4	943	13	BQ927245	BQ927245 AGENCOURT
25	703.6	24.8	758	12	B1917156	B1917156 AGENCOURT
26	701.4	24.7	919	13	BQ652596	BQ652596 AGENCOURT
27	697.8	24.6	905	13	BX390897	BX390897 BX390897
28	696.8	24.5	781	10	BQ676290	BQ676290 AGENCOURT
29	695.4	24.5	914	13	BQ957208	BQ957208 AGENCOURT
30	694.8	24.5	917	13	BQ509684	BQ509684 AGENCOURT
31	690.8	24.3	785	13	BQ053012	BQ053012 UI-M-FCO-
32	688.4	24.2	1061	12	BQ543298	BQ543298 AGENCOURT
33	677.2	23.8	704	9	AU131948	AU131948 AGENCOURT
34	674	23.7	828	14	CA151273	CA151273 UI-M-FW0-
35	674	23.7	912	12	B1685523	B1685523 603308994
36	665.8	23.4	754	14	CA131541	CA131541 UI-M-FW0-
37	663.6	23.4	750	13	BQ571495	BQ571495 UI-M-FCO-
38	663.2	23.3	788	14	CA181066	CA181066 UI-M-FW0-
39	661.8	23.3	908	13	BQ422607	BQ422607 AGENCOURT
40	654.2	23.0	744	12	B1184681	B1184681 UNL-P-FN-
41	653.2	23.0	767	13	BQ613510	BQ613510 UI-M-FW0-
42	651.2	22.9	745	14	CA451068	CA451068 UI-M-FW0-
43	648.4	22.8	847	9	AU129746	AU129746 AGENCOURT
44	643.8	22.7	944	13	BQ919072	BQ919072 AGENCOURT
45	639.2	22.5	755	14	CA151545	CA151545 UI-M-FW0-

## ALIGNMENTS

RESULT 1  
LOCUS BC000063  
DEFINITION Homo sapiens, Similar to SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1, clone IMAGE:3508933, mRNA.  
ACCESSION BC000063  
VERSION BC000063.1 GI:12652634  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3206)  
Strausberg, R.  
Direct Submission  
Submitted (03-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Institute for Systems Biology

http://www.systemsbio.org  
contact: amadan@systemsbiology.org  
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia  
Greene, Mark Ketterman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 7 Row: b Column: 12  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 4507082  
This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers  
1..3206  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3508933"  
/tissue\_type="Brain, neuroblastoma"  
/clone\_lib="NIH MGC 19"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"

BASE COUNT 738 a 913 c 783 g 772 t

ORIGIN

Query Match 85.9%; Score 2439.4; DB 11; Length 3206;  
Best Local Similarity 94.9%; Pred. No. 0;  
Matches 2669; Conservative 0; Mismatches 11; Indels 133; Gaps 9;

QY 160 GGGCGGGTTTCAGTCTGTGGTCCAAAGGGGGCGCGGAGCCCTCAGAGGGGGGGCG 219  
DB 9 GGGCGGGTTTCAGTCTGTGGTCCAAAGGGGGCGCGGAGCCCTCAGAGGGGGGGCG 68

QY 220 CGGCTGCTGGCTTGGGCCC-GGCGAACTCCGGGGCCCTCTGTGGGAATGGCGCGGCTC 278  
DB 69 CGGCTGCTGGCTTGGGCGGGGCGAACTCCGGGGCCCTCTGTGGGAATGGCGCGGCTC 128

QY 279 CGGGTCAAGGGGTGACCGCTCCCGATGCGCGAGCGGCTTATCCGAGACCAAGTATGT 338  
DB 129 CGGGTCAAGGGGTGACCGCTCCCGATGCGCGAGCGGCTTATCCGAGACCAAGTATGT 188

QY 339 TGCCAGGAGCGGAATGACACTCAGGACCTTCCATGGGACCCCTCGCTATGGGGGA 398  
DB 189 TGCCAGGAGCGGAATGACACTCAGGACCTTCCATGGGACCCCTCGCTATGGGGGA 248

QY 399 ACCCTTCAGTCCGACCTGGCTGGCCAGTCAGGATGGATCAGTCCCGCAAGACCTG 458  
DB 249 ACCCTTCAGTCCGACCTGGCTGGCCAGTCAGGATGGATCAGTCCCGCAAGACCTG 308

QY 459 CCCCTCAGCAGATCCAGAGGTCCAGCAGCGGCTCCAAATCGAAACCAATGCAA 518  
DB 309 CCCCTCAGCAGATCCAGAGGTCCAGCAGCGGCTCCAAATCGAAACCAATGCAA 368

QY 519 AGAAAAAGAGATGCTGACAAAATTTACCTCAAAGGATTCGTAACCTGATACAGAA 578  
DB 369 AGAAAAAGAGATGCTGACAAAATTTACCTCAAAGGATTCGTAACCTGATACAGAA 428

QY 579 CCCAGCCTATATGATCTTGGCTTTTGAAGGAAACTGACACAGACTATCATGAGGA 638  
DB 429 CCCAGCCTATATGATCTTGGCTTTTGAAGGAAACTGACACAGACTATCATGAGGA 488

QY 639 AACGGCTAGATATCAAGAGGCTTTGAAACGTCCCATTAAGCAAAACCGAAGCTGCAA 698  
DB 489 AACGGCTAGATATCAAGAGGCTTTGAAACGTCCCATTAAGCAAAACCGAAGCTGCAA 548

QY 699 TTTTCATTTCAACTTCCGCTTAAGTACAGTCCGAGATTCAGCTTGTCCAAAT 758  
DB 549 TTTTCATTTCAACTTCCGCTTAAGTACAGTCCGAGATTCAGCTTGTCCAAAT 608

QY 759 TGGCTTCTGGAGCTTCCGGGTAGAGGACGGCTCCTGAGGATTCAGCCTTGTCCAAAT 818  
DB 609 TGGCTTCTGGAGCTTCCGGGTAGAGGACGGCTCCTGAGGATTCAGCCTTGTCCAAAT 668

QY 819 ATGATGCCCAATAACAAAGAGGAAGTTCTCTTCC-TTCTTTAAGTCTCTGGTATTGAA 877

DB 669 ATGATGCCCAATAACAAAGAGGAAGTTCTCTCTCTTTTAAAGTCTCTGGTATTGAA 728  
QY 878 CTGACAAAGACCTGTATGGCCAGACAAACCATCTGTGTAGATGGCCAGACGCGCACT 937  
DB 729 CTGACAAAGACCTGTATGGCCAGACAAACCATCTGTGTAGATGGCCAGACGCGCACT 788  
QY 938 ACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCGAGGAGATGTGAATGTACGGTACT 997  
DB 789 ACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCGGAGACGCTGAATGTACGGTACT 848  
QY 998 GTCTACTGATGCTGGATTACAGCCTCCAGAGTTTAAATTTAGACCCCGCTAGCTCGA 1057  
DB 849 GTCTACTGATGCTGGATTACAGCCTCCAGAGTTTAAATTTAGACCCCGCTAGCTCGA 908  
QY 1058 CTCTGGGCTATCCATACCCAGACTCGTCCAGTGATCATCAAAGCACTGTGCAATATATT 1117  
DB 909 CTCTGGGCTATCCATACCCAGACTCGTCCAGTGATCATCAAAGCACTGTGCAATATATT 968  
QY 1118 AAGACACATAAGCTCCAGGACCCCTCAGAGCGGAGTTTGTCTATCTGTGACAAAGTACTCG 1177  
DB 969 AAGACACATAAGCTCCAGGACCCCTCAGAGCGGAGTTTGTCTATCTGTGACAAAGTACTCG 1028  
QY 1178 CAGCAGATCTTTGAGACTCAACGATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCC 1237  
DB 1029 CAGCAGATCTTTGAGACTCAACGATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCC 1088  
QY 1238 TTGCTTATGCCACAGAACCTATCATCATTAATCATGTCACTCAGTGTGTGACCCCAATGAT 1297  
DB 1089 TTGCTTATGCCACAGAACCTATCATCATTAATCATGTCACTCAGTGTGTGACCCCAATGAT 1148  
QY 1298 CAGAAAAAGACAGCTTGTATGACATTTGTAAGTGGATGACACCTTTGAAGACCCAG 1357  
DB 1149 CAGAAAAAGACAGCTTGTATGACATTTGTAAGTGGATGACACCTTTGAAGACCCAG 1208  
QY 1358 ATGAATTTCTTTCTGCTGTCCA CTGCCAGCCAA CAGAGATTTGCTACTCTAGACAA --- 1413  
DB 1209 ATGAATTTCTTTCTGCTGTCCA CTGCCAGCCAA CAGAGATTTGCTACTCTAGACAA CAG 1268  
QY 1414 ----- 1413  
DB 1269 ATCCATGAGACAATAGAAACCATCAACAGCTGAGACTCAGCGGAGTTCTATGCTGAGC 1328  
QY 1414 -----C 1414  
DB 1329 TTTGCCAGAGACCTCAGGGTTTTCATCAATGACTGGCTTCAGTCCAGTGCGAGGACCTC 1388  
QY 1415 AAGCAATGACTGATGTGGGTAACTCCAGAGAGGAGCGCCGAGCTGAGTGTCTACTTTC 1474  
DB 1389 AAGCAATGACTGATGTGGGTAACTCCAGAGAGGAGCGCCGAGCTGAGTGTCTACTTTC 1448  
QY 1475 CAGCCCTGGGCTCAGGAGGCTGTGTGCGGATCTTCTACTCCAAGGTGCGAGCAGACGA 1534  
DB 1449 CAGCCCTGGGCTCAGGAGGCTGTGTGCGGATCTTCTACTCCAAGGTGCGAGCAGACGA 1508  
QY 1535 CAAGAAATTAGAGCAAGCCCTGGGAATCCGGAATACATAGGCGCTCTCCCAAGGCTGAT 1594  
DB 1509 CAAGAAATTAGAGCAAGCCCTGGGAATCCGGAATACATAGGCGCTCTCCCAAGGCTGAT 1568  
QY 1595 TCAGCTGACCAATTTCTTGATTTGGGCGCTGTGTGCTGCTCATAGTATCTGCTTGG 1654  
DB 1569 TCAGCTGACCAATTTCTTGATTTGGGCGCTGTGTGCTGCTCATAGTATCTGCTTGG 1628  
QY 1655 TCTTGTCTGGGCGCTCCAGGGATGCTGTGTGTTCAAGGACAGACAGCAATCAAGAGG 1714  
DB 1629 TCTTGTCTGGGCGCTCCAGGGATGCTGTGTGTTCAAGGACAGACAGCAATCAAGAGG 1689  
QY 1715 GTCTCACAAGACACTGTATTCTCTTTTACCCCTATCTCTCCACCCAGCTTCC 1774  
DB 1689 GTCTCACAAGACACTGTATTCTCTTTTACCCCTATCTCTCTCCACCCAGCTTCC 1748  
QY 1775 CTTTGGCCCAAAAGTTCCCATGTGCTGTACCTCCCTGCTCTACATAGGACCTCTAG 1834



```
QY 393 GGGGAACCCCTTCAGTCCGACCTCGCTGGCCAGCTCAGGAGTGGATCAGTCCCGCAAGA 452
Db 361 GGGGAACCCCTTCAGTCCGACCTGGCTGGCCAGTCCAGGATGGATCAGTCCCGCAAGA 420
QY 453 GACCTGCCCTCAGCAGATCCAGCAGGTCCAGCAGCGGTCCAAATCGAAACCAACA 512
Db 421 GACCTGCCCTCAGCAGATCCAGCAGGTCCAGCAGCGGTCCAAATCGAAACCAACA 480
QY 513 ATGCAAGAGAAAAGAGTGGCTGACAAAATTTCTACCTCAAGAGATTCTGTAACCTGGTAC 572
Db 481 ATGCAAGAGAAAAGAGTGGCTGACAAAATTTCTACCTCAAGAGATTCTGTAACCTGGTAC 540
QY 573 CAGAAATCCAGGCTATATGGATCTCTTGGCTTTTGAAGGAAACTGGACAGACTATCA 632
Db 541 CAGAAATCCAGGCTATATGGATCTCTTGGCTTTTGAAGGAAACTGGACAGACTATCA 600
QY 633 TGAGAAACGGCTAGATATCCAGAGGCTTGAACGTCCTCCATTAAGCAAAAACGGAAGC 692
Db 601 TGAGAAACGGCTAGATATCCAGAGGCTTGAACGTCCTCCATTAAGCAAAAACGGAAGC 660
QY 693 TCGGAATTTTCAATTTTCAACACTTTCAATCCGGCTAAGTCAAGTCCGAGGATGGGAAG 752
Db 661 TCGGAATTTTCAATTTTCAACACTTTCAATCCGGCTAAGTCAAGTCCGAGGATGGGAAG 720
QY 753 GGACGGTGGCTTCTCGGAGCTTCGGGTAGAGAGCGCTCTCGGAGGATTCAGCTTTGT 812
Db 721 GGACGGTGGCTTCTCGGAGCTTCGGGTAGAGAGCGCTCTCGGAGGATTCAGCTTTGT 780
QY 813 CCAATATGATGCCACTTAACAAAGAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 872
Db 781 CCAATATGATGCCACTTAACAAAGAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
QY 873 TTGAATCGACAAAGACCTGTATGGCCAGACAAACCATCTGGTAGAATGGCACAGGACCG 932
Db 841 TTGAATCGACAAAGACCTGTATGGCCAGACAAACCATCTGGTAGAATGGCACAGGACCG 900
QY 933 CCACCTACCCAGGAGCCGATGGCTT 957
Db 901 CCACCTACCCAGGAGCCGATGGCTT 925

RESULT 3
BX386559 933 bp mRNA linear EST 08-MAY-2003
LOCUS
DEFINITION
BX386559 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens cDNA clone CS0DJ005YB22 5-PRIME, mRNA sequence.
ACCESSION
BX386559
VERSION
BX386559.1 GI:30455451
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 933)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6875.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSIAJ002B120P1&cluster=6875.f. Contact :
Feng liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSIAJ002B120P1.
Location/Qualifiers
1..933
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
BASE COUNT 230 a 245 c 275 g 177 t 6 others
ORIGIN
Query Match 29.8%; Score 847.8; DB 13; Length 933;
Best Local Similarity 96.7%; Pred. No. 3.5e-214;
Matches 869; Conservative 6; Mismatches 23; Indels 1; Gaps 1;
QY 112 CCGGTTCTTTGTGGCGTGCATCGGCGCTCCGGGAAGATGGCGGCCGGCGGGTTTCC 171
Db 35 CAGGCTGGTACCGGTCGGGAATTCCTGGGATCGGGAAGATGGCGGCCGGCGGGTTTCC 94
QY 172 AGTCTGTGGCTCCAAAGCGCGCGCGAGCCTCAGGAGGGCGGGCGCGCTCGCTCGCT 231
Db 95 AGTCTGTGGCTCCAAAGCGCGCGCGAGCCTCAGGAGGGCGGGCGCGCTCGCTCGCT 154
QY 232 TGGCCCTC-GGCGGAATCTCCGGGGCTCTCTGTGCGAATGGGCCCGCGCTCCGGGTCAAGGGC 290
Db 155 TGGSCCCCGGCGGAATCTCCGGGGCTCTCTGTGCGAATGGGCCCGCGCTCCGGGTCAAGGGC 214
QY 291 TGTAACCTCTCCCGATCCCGAGCGGCTATCGAGACAGAGTATGTCGAGGAGCC 350
Db 215 TGTACCTCTCCCGATCCCGAGCGGCTATCGARACAGAGTATGTCGAGGAGCC 274
QY 351 GAATGACACCTCAGGAGACCTTCATGGGACCCCTGCTATGGGGAAACCTTCAGTCC 410
Db 275 GAATGACACCTCAGGAGACCTTCATGGGACCCCTGCTATGGGGAAACCTTCAGTCC 334
QY 411 GACCTGGCTGGCCAGTCAAGGATGGATCAGTCCCGCAAGAGACCTTCGCCCTCAGCAGA 470
Db 335 GACCTGGCTGGCCAGTCAAGGATGGATCAGTCCCGCAAGARACCTTCGCCCTCAGCAGA 394
QY 471 TCCAGAGGTCAGCAGCAGCGGTCCAAATCGAAACCAATCGAAAGAAAGAAAGA 530
Db 395 TCCAGAGGTCAGCAGCAGCGGTCCAAATCGAAATCGAAACCAATCGAAAGAAAGAAAGA 454
QY 531 TGGCTGACAAATTTCTACCTCAAGAGTTCGTGAACCTGGTACCAAGATCCAGGCTTATA 590
Db 455 TGGCTGACAAATTTCTACCTCAAGAGTTCGTGAACCTGGTACCAAGATCCAGGCTTATA 514
QY 591 TGGATCTCTTGGCTTTTGAAGGAAACTGGACCAAGACTATCATGAGGAAACGGCTAGATA 650
Db 515 TGGATCTCTTGGCTTTTGAAGGAAACTGGACCAAGACTATCATGAGGAAACGGCTAGATA 574
QY 651 TCCAAAGAGGCTTGAAACGTCCTCAATTAAGCAAAAACGGAAGCTCGAATTTTCTATA 710
Db 575 TCCAAAGAGGCTTGAAACGTCCTCAATTAAGCAAAAACGGAAGCTCGAATTTTCTATA 634
QY 711 ACCTTTTCAATCCGGCTAAGTCCAGATCCGAGGATGGGAGGAGCGTTCCTCTGGG 770
Db 635 ACCTTTTCAATCCGGCTAAGTCCAGATCCGAGGATGGGAGGAGCGTTCCTCTGGG 694
QY 771 AGCTTCGGGTAGAGGAGCGCTCCTCGAGGATTCAGGCTTTGTCCAAATATGATGCCACTA 830
Db 695 AGCTTCGGGTAGAGGAGCGCTCCTCGAGGATTCAGGCTTTGTCCAAATATGATGCCACTA 754
QY 831 AACAAAAGAGAGTTCCTCTCTCTTTTAAAGTCTTTGGTGAATGAACGAGCAAGAGCC 890
Db 755 AACAAAAGAGAGTTCCTCTCTCTTTTAAAGTCTTTGGTGAATGAACGAGCAAGAGCC 814
QY 891 TGTATGGCCAGACCAACCATCTGTGTAATGGACAGGAGCCGACCTACCCAGGAGCCG 950
Db 815 TGTATGGCCAGACCAACCATCTGTGTAATGGACAGGAGCCGACCTACCCAGGAGCCG 874
```

```
/db_xref="taxon:9606"
/clone="CS0DJ005YB22"
/cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
/cell_line="JURKAT"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
10-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
```

Db	369	CCGCGAAGAAACCTGCCCCCTCAGAGATCCAGCAGGCTCCAGCAGCGCGGTCCAAATC	428
QY	504	GAACCAACAATGCAAGAAAGAAAGATGCGCTGACAAATTTCTA	563
Db	429	GAACCAACAATGCAAGAAAGAAAGATGCGCTGACAAATTTCTA	488
QY	564	AACCTGTACCAAGATCCAGCGCTATATGATCTCTTGGCTTTTGA	623
Db	489	AACCTGTACCAAGATCCAGCGCTATATGATCTCTTGGCTTTTGA	548
QY	624	AGACTATCATAGAGAAACGGCTAGATATCCAAAGAGCGCTTGA	683
Db	549	AGACTATCATAGAGAAACGGCTAGATATCCAAAGAGCGCTTGA	608
QY	684	AACGGAAGCTGCGAATTTT-CATTCTAACACTTTCATCCGCTAA	742
Db	609	AACGGAAGCTGCGAATTTTACATTCTTAACACTTTCATCCGCTAA	668
QY	743	GATGGGGGAAGGACGGTGGCTTCTGGGAGCTTGGGTAGAAAGA	802
Db	669	GATGGGGGAAGGACGGTGGCTTCTGGGAGCTTGGGTAGAAAGA	728
QY	803	TCAGCTTTGTCCAAATATGATGCCACTTAAACAAAGAGAAAGT	861
Db	729	TCAGCTTTGTCCAAATATGATGCCACTTAAACAAAGAGAAAGT	788
QY	862	GTCCTTGGTATGATGAATGAGCAAAAGACCTGTATGGCCAGACA	921
Db	789	GTCCTTGGTATGATGAATGAGCAAAAGACCTGTATGGCCAGACA	848
QY	922	GCACAGACCGGCCACTCTCCAGAGAGACCATGCTTCCAGCTGA	981
Db	849	GCACAGACCGGCCACTCTCCAGAGAGACCATGCTTCCAGCTGA	908
QY	982	GAATGTACGGGTACTCTCTCCACTGATGCGGATTAACAGCGCT	1041
Db	909	GAATGTACGGGTACTCTCTCCACTGATGCGGATTAACAGCGCT	967
QY	1042	CCCCC 1046	
Db	968	CCCCC 972	
RESULT 5	B0929257		
LOCUS	B0929257		
DEFINITION	AGENCOURT 10036211 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6482661		
ACCESSION	B0929257		
VERSION	B0929257.1		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	B0929257		
AUTHORS	Eukaryotic Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	1 (bases 1 to 907)		
COMMENT	NIH-MGC http://mgs.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. Email: gspires-remail.nih.gov Tissue Procurement: DCTD/DTF cDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: L1CM2665 row: p column: 22 High quality sequence stop: 705. Location/Qualifiers 1..907		

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6482661"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/clone="organ: prostate; Vector: pOTB7; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

```

```

BASE COUNT      255 a 231 c 223 g 198 t
ORIGIN
Query Match      29.3%; Score 833.6; DB 13; Length 907;
Best Local Similarity 97.6%; Pred. No. 2.1e-210;
Matches 867; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

QY 439 TCAGTCCCGCAAGAGACCTGCCCTCAGCAGATCCAGCAGGTCCAGAGCGGTCCA 498
Db 1 TCAGTCCCGCAAGAGACCTGCCCTCAGCAGATCCAGCAGGTCCAGAGCGGTCCA 60

QY 499 AAATCGAAACCAATGCAAGAAAGAAAGATGGCTGACAAATTCCTACCTCAAAGGAT 558
Db 61 AAATCGAAACCAATGCAAGAAAGAAAGATGGCTGACAAATTCCTACCTCAAAGGAT 120

QY 559 TCGTGAACCTGTACCAGAAATCCAGCGCTATATGGATCTCTTGGCTTTGAAAGGAAACT 618
Db 121 TCGTGAACCTGTACCAGAAATCCAGCGCTATATGGATCTCTTGGCTTTGAAAGGAAACT 180

QY 619 GGACAGACTATCATGAGAAACGGCTAGATATCCAAAGGCGCTTGAACGTCCTCCATTAA 678
Db 181 GGACAGACTATCATGAGAAACGGCTAGATATCCAAAGGCGCTTGAACGTCCTCCATTAA 240

QY 679 GCARAAACGGAGCTGCCAATTTTCATTTCTAACACATTTCAATCCGGCTAGTCAGATGC 738
Db 241 GCAAAACGGAGCTGCCAATTTTCATTTCTAACACATTTCAATCCGGCTAGTCAGATGC 300

QY 739 CCAGATGGGGAAGGACGGTGGCTTCCTGGAGCTTCGGGTAGAAAGACGGCTCCTGGA 798
Db 301 CCAGATGGGGAAGGACGGTGGCTTCCTGGAGCTTCGGGTAGAAAGACGGCTCCTGGA 360

QY 799 GAATTCAGCTTGTCCAAATATGATGCGCACTAAACAAAGAGGAAGTTCTCTTCTCTTT 858
Db 361 GAATTCAGCTTGTCCAAATATGATGCGCACTAAACAAAGAGGAAGTTCTCTTCTCTTT 420

QY 859 TAACTCCTTGGTATGTAACCTGGACAAAGACCTGTATGGCCAGACACCACTCTGGTGA 918
Db 421 TAACTCCTTGGTATGTAACCTGGACAAAGACCTGTATGGCCAGACACCACTCTGGTGA 480

QY 919 ATGGCAGAGGACCGCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAGA 978
Db 481 ATGGCAGAGGACCGCCACTACCCAGGAGACCGATGGCTTCCAGGTGAAGCGGCCAGGAGA 540

QY 979 TGTGAATGTACGGTGTACTGTCTTACTGATGTGTGATTAACAGCTCCCTCCAGTTTAAATT 1038
Db 541 COTGAATGTACGGTGTACTGTCTTACTGATGTGTGATTAACAGCTCCCTCCAGTTTAAATT 600

QY 1039 AGACCCCGCTAGCTCCAGCTCCTGGGATCCATACCCAGACTCGTCAGGTGATCATCCA 1098
Db 601 AGACCCCGCTAGCTCCAGCTCCTGGGATCCATACCCAGACTCGTCAGGTGATCATCCA 660

QY 1099 AGCACTGTGGCAATATATTAGACACATAAGCTCCAGGACCTCCACGA -GGGGAGTTTG 1157
Db 661 AGCACTGTGGCAATATATTAGACACATAAGCTCCAGGACCTCCACGAAGGGAGTTTG 720

QY 1158 TCATCTGTGACAAGTACCTGACAGAGATCTTTGAGACTCAACGATATGAATTTTCAGAGA 1217
Db 721 TCATCTGTGACAAGTACCTGACAGAGATCTTTGAGCTCCAGTATGAAGTTTTCAGAGA 780

```

```

QY 1218 TCCTCAGCGCTCCATGCTTGTATGCGCACCAACCTATCATTAATCAT-GTC 1276
Db 781 TCCTCAGCGCTCCATGCTTGTATGCGCACCAACCTATCATTAATCATCTGGC 840

QY 1277 ATCAGTTTGACCGAATGATCAGAAAAGACAGCTTTGTTATGACATT 1324
Db 841 ATCAGTGGTGACCGCATGATCAAAAAAGACCGCTTGGTATGACACT 888

```

## RESULT 6

## AL521992

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## UNPUBLISHED

## CONTACT

## GENOSCOPE

## BP 191 91006 EVRY cedex - France

## Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr

## Library was constructed by Life Technologies, a division of

## Invitrogen. This sequence belongs to sequence cluster 6875.f For

## more information about this cluster, see

## http://www.genoscope.cns.fr/

## cgi-bin/cluster.cgi?seq=CS0DB007AG01QPI&amp;cluster=6875.f. Contact :

## Feng Liang Email : fliang@lifetech.com URL :

## http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

## Faraday Avenue Genoscope sequence ID : CS0DB007AG01QPI.

## Location/Qualifiers

## 1. .970

## /organism="Homo sapiens"

## /mol\_type="mRNA"

## /db\_xref="taxon:9606"

## /clone="CS0DB007YM01"

## /tissue\_type="NEUROBLASTOMA COT 10-NORMALIZED"

## /clone\_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"

## /notes="1st strand cDNA was primed with a NotI-oligo(dT)

## primer. Five prime end enriched, double-strand cDNA was

## digested with Not I and cloned into the Not I and EcoR V

## sites of the pCMVSPORT 6 vector. Library was normalized."

## BASE COUNT

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## Conservative

## Mismatches

## Indels

## Gaps

## Length

## DB 9;

## Pred. No. 3.7e-208;

## 0; Mismatches

## 3; Indels

## 8; Gaps

## 6;

## Length

## 970;

## DB 9;

## Pred. No. 3.7e-208;

## 0; Mismatches

## 3; Indels

## 8; Gaps

## 6;

## Length

## 970;

## DB 9;









REFERENCE	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	
2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
TITLE	Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to
MEDLINE	prepare full-length cDNA libraries for rapid discovery of new genes
PUBMED	Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE	
3	
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
JOURNAL	Kono, H., Akiyama, Y., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M.,
MEDLINE	Sun, N., Ishii, Y., Nakamura, S., Hazana, M., Nishie, T., Harada, A.,
PUBMED	Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K.,
REFERENCE	Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
4	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,
TITLE	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
JOURNAL	RIKEN integrated sequence analysis (RISA) system--384-format
MEDLINE	sequencing pipeline with 384 multiplexed sequencing
PUBMED	Genome Res. 10 (11), 1757-1771 (2000)
REFERENCE	
5	
TITLE	Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,
JOURNAL	Arkawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,
MEDLINE	Aizawa, K., Iwawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,
PUBMED	Salto, T., Okazaki, Y., Gotohori, T., Bono, H., Kasukawa, T., Salto, R.,
REFERENCE	Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,
6	Flatschman, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,
7	Kuehl, P., Lewis, S., Matsuo, Y., Nakaido, I., Pesole, G.,
TITLE	Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M.,
JOURNAL	Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,
MEDLINE	Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,
PUBMED	Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bull, C.,
REFERENCE	Flatcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,
8	Hofmann, M., Hume, D.A., Kamlay, M., Lee, N.H., Lyons, P.,
9	Matchionni, L., Mashima, J., Mazzarelli, J., Mombereis, P., Nordone, P.,
TITLE	Ring, B., Ringuet, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,
JOURNAL	Sato, K., Schombach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,
MEDLINE	Toyo-Oka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilmberg, L.,
PUBMED	Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S.
REFERENCE	and Hayashizaki, Y.
10	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409 (6821), 685-690 (2001)
MEDLINE	21085660
PUBMED	11217851
REFERENCE	
11	
TITLE	The FANTOM Consortium and the RIKEN Genome Exploration Research
JOURNAL	Group Phase I & II team.
MEDLINE	Analysis of the mouse transcriptome based on functional annotation
PUBMED	of 60,770 full-length cDNAs
REFERENCE	Nature 420, 563-573 (2002)
12	(bases 1 to 2476)
13	
TITLE	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
JOURNAL	Arkawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
MEDLINE	Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
PUBMED	Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
REFERENCE	Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M.,
14	Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
15	Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
TITLE	Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
JOURNAL	Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
MEDLINE	Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
PUBMED	Tejima, Y., Toya, T., Yamamura, T., Yamanka, I., Yasunishi, A.,
REFERENCE	Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
16	Direct Submission
17	Submitted (16-APR-2002) Yoshinide Hayashizaki, The Institute of
18	Physical and Chemical Research (RIKEN), Laboratory for Genome

COMMENT  
Exploration Research Group RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gs.c.riken.go.jp,  
URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.go.jp/>  
URL: <http://fantom.gsc.riken.go.jp/>.  
Location/Qualifiers

FEATURES  
source  
1 . 2476  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="db\_xref="PANTOM,DB:0710008A09"  
/db\_xref="db\_xref="MGI:190344"  
/db\_xref="db\_xref="taxon:10090"  
/clone="0710008A09"  
/sex="male"  
/tissue\_type="brain"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
1 . 920  
/note="unnamed protein product; SWI/SNF related, matrix  
associated, actin dependent regulator of chromatin,  
subfamily d, member 1 (MGI:MGI:1933623)  
putative"  
/codon\_start=3  
/protein\_id="BAC35856.1"  
/db\_xref="db\_xref="GI:26344411"  
/db\_xref="db\_xref="MGI:1933623"  
/translation="VEGRILIEEALSKYDAITKOKRSPFKSLVITELDKLVGPDNHH  
LVENHRRTATTOETDGFQVYKRPBGVNNACTVLMLDYPPOPKDPRLARLIGITQTR  
PVIIQALWQYIKTHKLDPRHREDFVLCCKYLQOIPEEQRMKFSIIPORLHALMPPE  
IINIHVLSVDNDKQKTCACVADIVEDVDLKTQNSFLSLSTAQOEIATDNKIHETI  
ETIOLKTRFPMLSFARDPQGFINDMLQSCRDILKMTDVGVPBEERRAEFYFQW  
AQEVCRFYFSKVQGRQRELEQLGINT"  
2463 . 2468  
polya\_signal  
/note="putative"  
2476  
polya\_site  
/note="putative"  
BASE COUNT 513 a 743 c 622 g 598 t  
ORIGIN  
Query Match 28.0%; Score 795.6; DB 11; Length 2476;  
Best Local Similarity 69.6%; Pred. No. 3.8e-200;  
Matches 1530; Conservative 0; Mismatches 409; Indels 259; Gaps 23;  
OY 777 GGGTAGAAGACGGCTCTGGAGATTACGCTTGCCAAATATGATGCCCTAAACAA 836  
DB 1 GGGTAGAAGCGGGATCTCGAGGAACGGCTTGCCAAATATGAGCCACCAAGCAA 60  
OY 837 AGAGAGATTCTCTCTCTTTAAGTCCTTGATGTGAATGACCAAGAACCTGTATG 896  
DB 61 AGAAGAGTCTCTCTCTTTTAAAGCTCTTGATGCAACTGGACAAAGACCTCATG 120  
OY 897 GGCCAGAACCAATCTGTATGATGCGACAGAGACCGCACTACCCAGAGACGATGCT 956  
DB 121 GCCCAGAACCAATCTGTATGAAAGGACAGAGACCGCACTACCCAGAGACGATGCT 180  
OY 957 TCCAGGTGAAGCGGCGAGAGATGTAATGATAGCGTACTGTCCTCATGATGCGAT 1016  
DB 181 TCCAGGTGAAGCGGCGAGAGATGTAATGATAGCGTACTGTCCTCATGATGCGAT 240  
OY 1017 ACCAGCTTCCCAATTAAATTAGACCCCGCGCTAGCTGACTCTGGGATCCATACC 1076  
DB 241 ACCAGCCCCCCAGTTAAATTAGACCTCGCTGCGCTCTTGCGATCATACCC 300  
OY 1077 AGACTCGTCAGTATCATCCAGCACTGGCGCATATATTAAAGACATTAAGTCCAG 1136



Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHC2536 row: j column: 20  
 High quality sequence start: 36  
 High quality sequence stop: 694.  
 Location/Qualifiers  
 1. 972  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6340339"  
 /issue\_type="neuroblastoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 47"  
 /note="Organ: Brain; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCAAGAG(G). Size selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH MGC Library."

BASE COUNT 262 a 246 c 240 g 223 t 1 others

Query Match 27.4% Score 779.8; DB 13; Length 972;  
 Best Local Similarity 98.3%; Pred. No. 4.5e-196;  
 Matches 788; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

608 GAAAGAACTGGACAGACTATCATAGAGAAACGGCTAGATATCCAGAGCCCTTGAA 667  
 16 GAGGGAACTGACACAGACTTCTCGAAGAAACGGCTAGATATCCAGAGCCCTTGAA 75  
 668 CGTCCATTAAAGCAAAACGGAAGCTGCGAATTTTCAATTTCAACTTCAATCCGCT 727  
 76 CGTCCATTAAAGCAAAACGGAAGCTGCGAATTTTCAATTTCAACTTCAATCCGCT 135  
 728 AAGTCAGATGCCAGATGGGAGGAGCGTGGCTTCTGGAGAGCTTGGGATGAAGA 787  
 136 AAGTCAGATGCCAGATGGGAGGAGCGTGGCTTCTGGAGAGCTTGGGATGAAGA 195  
 788 CGGCTCTGGAGATTCAGGCTTGTCCAAATATGATGCCACTAAACAAAGAGAGATT 847  
 196 CGGCTCTGGAGATTCAGGCTTGTCCAAATATGATGCCACTAAACAAAGAGAGATT 255  
 848 TCTTCT-TTCTTTAAGTCTTGTGATGAATGGAACCAAGAGCTGTATGGCCAGACA 906  
 256 TCTTCTTTTAAAGTCTTGTGATGAATGGAACCAAGAGCTGTATGGCCAGACA 315  
 907 CCATCTGTGATGATGGACAGGACCGCCACTACCCAGAGAGCCGATGGCTTCAAGTGA 966  
 316 CCATCTGTGATGATGGACAGGACCGCCACTACCCAGAGAGCCGATGGCTTCAAGTGA 375  
 967 GGGGCGAGGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1026  
 376 GGGGCGAGGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 435  
 1027 CCAGTTAAATTAGACCCCGCTAGCTGACTCTGGGATCCATACCCAGAGCTGCTC 1086  
 436 CCAGTTAAATTAGACCCCGCTAGCTGACTCTGGGATCCATACCCAGAGCTGCTC 495  
 1087 AGGATCATCAAGCACTGTGGCAATATATTAAGACACATTAAGCTCCAGAGCCCTCACA 1146  
 496 AGGATCATCAAGCACTGTGGCAATATATTAAGACACATTAAGCTCCAGAGCCCTCACA 555  
 1147 GCGGAGTTTGTATGTGACAAAGTACCTGACAGAGATCTTGAAGCTCAAGCTATGAA 1206

Db 556 GCGGAGTTTGTATGTGACAAAGTACCTGACAGAGATCTTGAAGCTCAAGTATGAA 615  
 QY 1207 GTTTTCAGAGATCCCTCAGCGGCTCCATGCTTGTATGACACAGAACCTATCATCAT 1266  
 Db 616 GTTTTCAGAGATCCCTCAGCGGCTCCATGCTTGTATGACACAGAACCTATCATCAT 675  
 QY 1267 TAATCATGTCATCAGTGTGACCCGCAATGATCAGAAAAAGACAGCTTGTATGATGATGA 1326  
 Db 676 TAATCATGTCATCAGTGTGACCCGCAATGATCAGAAAAAGACAGCTTGTATGATGATGA 735  
 QY 1327 TGTGAGTGTGATGACACCTTGAAGACCCAGATGATTTCTTGTCTGTCTCAGTGGCAG 1386  
 Db 736 TGTGAGTGTGATGACACCTTGAAGACCCAGATGATTTCTTGTCTGTCTCAGTGGCAG 795  
 QY 1387 CCAACAGAGATTTGCTACTCTGACACACAGA 1418  
 Db 796 CCAACAGAGATTTGCTACTCTGACACACAGA 827

RESULT 11  
 BU931457 872 bp mRNA linear EST 18-OCT-2002  
 LOCUS AGENCOURT\_10478213 NIH\_MGC\_109 Homo sapiens cDNA clone  
 DEFINITION IMAGE:6671629 5', mRNA sequence.  
 BU931457  
 ACCESSION BU931457.1 GI:24120276  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 872)  
 NIH-MGC <http://mgi.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-rc@mail.nih.gov](mailto:cgabbs-rc@mail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LHC2950 row: n column: 13  
 High quality sequence stop: 646.  
 Location/Qualifiers  
 1. 872  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6671629"  
 /issue\_type="teratocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 109"  
 /note="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:  
 XhoI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAAGAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH MGC Library."

BASE COUNT 209 a 238 c 261 g 164 t

Query Match 27.2% Score 772.4; DB 13; Length 872;  
 Best Local Similarity 98.7%; Pred. No. 4.1e-194;  
 Matches 831; Conservative 0; Mismatches 6; Indels 5; Gaps 5;

139 GCTCCGGAAGATGCGCGCGCGGCTTTCAGTCTGTGGCTTCAAGCGGCGCGCG 198  
 1 GCTCCGGAAGATGCGCGCGCGGCTTTCAGTCTGTGGCTTCAAGCGGCGCGCGCG 60

QY	199	GAGCCTCAGGAGGGCGGGCTGCTGCTTGGGCC-GGCGAACTCCGGGGCCTC	257
Db	61	GAGCCTCAGGAGGGCGGGCTGCTGCTTGGGCCGGCGGAACTCCGGGGCCTC	120
QY	258	CTGTGCGAATGGCCCGGCTCCGGGTCAAGGGCTATACGGTCCCGCATGCCCGAGGG	317
Db	121	CTGTGCGAATGGCCCGGCTCCGGGTCAAGGGCTATACGGTCCCGCATGCCCGAGGG	180
QY	318	CTATCCGAGACAGGTATGTTGCCAGGCGCCGAATGACACTCAGGAGCTTCCATGG	377
Db	181	CTATCCGAGACAGGTATGTTGCCAGGCGCCGAATGACACTCAGGAGCTTCCATGG	240
QY	378	GACCCCTTGGCTATGGGGGAAACCCCTTCAGTCCGACCTGGCTGCCCGCAGTCAGGGATGG	437
Db	241	GACCCCTTGGCTATGGGGGAAACCCCTTCAGTCCGACCTGGCTGCCCGCAGTCAGGGATGG	300
QY	438	ATCAGTCCGCAAGAGACTGCGCCCTCAGCAGATCCAGAGTCCAGCAGAGCGCGTCC	497
Db	301	ATCAGTCCGCAAGAGACTGCGCCCTCAGCAGATCCAGAGTCCAGCAGAGCGCGTCC	360
QY	498	AAATCGAAACACCAATCCAAAGAAAGAGATGGCTGACAAATTTCTACTCAAAGGA	557
Db	361	AAATCGAAACACCAATCCAAAGAAAGAGATGGCTGACAAATTTCTACTCAAAGGA	420
QY	558	TTCTGTGAATCGTACCAAGATCCAGGCTATATGGATCTCTTGGCTTTTGAAGGAAAC	617
Db	421	TTCTGTGAATCGTACCAAGATCCAGGCTATATGGATCTCTTGGCTTTTGAAGGAAAC	480
QY	618	TGGACACAGATCATAGAGGAAACCGGCTAGATATCCAGAGGCTTTGAAACGTCCCATTA	677
Db	481	TGGACACAGATCATATGAGGAAACCGGCTAGATATCCAGAGGCTTTGAAACGTCCCATTA	540
QY	678	AGCAAAACGGAAGCTCGAATTTCAATTTCTTAACACTTCAATCCGCTAAGTCAGATG	737
Db	541	AGCAAAACGGAAGCTCGAATTTCAATTTCTTAACACTTCAATCCGCTAAGTCAGATG	600
QY	738	CCGAGGATGGGAAGGAGCGGTGGCTTCTCGGAGCTTCGGGTGAAGAGCGGCTCCTGG	797
Db	601	CCGAGGATGGGAAGGAGCGGTGGCTTCTCGGAGCTTCGGGTGAAGAGCGGCTCCTGG	660
QY	798	AGGATTCAGCCTTTGCAAAATATGATGCACTAAACAAAGAGGAAGTTCTCTTCC-TTC	856
Db	661	AGGATTCAGCCTTTGCAAAATATGATGCACTAAACAAAGAGGAAGTTCTCTCTTTT	720
QY	857	TTTAAAGTCTTGGTGAATGAACTGGACCAAGACCTGTATGGCCAGACCAATCTGGTA	916
Db	721	TTTAAAGTCTTGGTGAATGAACTGGACCAAGACCTGTATGGCCAGACCAATCTGGTA	780
QY	917	GAATGGCACAGGA-CCGCCACTTACCC-AGGAGACCGAT-GGCTTCCAGGTGAAGCGGCA	973
Db	781	GAATGGCACAGGACCGGCCACTACCCAGGAGACCGATGGGCTTTCAGGTGAAGCGGCC	840
QY	974	GG 975	
Db	841	GG 842	

RESULT 12  
BM457092  
LOCUS  
DEFINITION BM457092 1003 bp mRNA linear EST 05-FEB-2002  
AGENCY 6406599 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:558337  
5' mRNA sequence.  
ACCESSION BM457092  
VERSION BM457092.1 GI:18506132  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1003)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL COMMENT

Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-rc@mail.nih.gov](mailto:cgabbs-rc@mail.nih.gov)  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM12346 row: e column: 02  
High quality sequence stop: 647.  
Location/Qualifiers  
1. 1003  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:558337"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_92"  
/notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 Kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC Library."  
BASE COUNT 190 a 308 c 236 g 269 t

FEATURES source

QY	1490	GAGCCTGTGGCGGATACATAGGGCTCTCCACAGCCCTGATTCGACTGCACCAATT	1609
Db	1	GAGCCTGTGGCGGATACATAGGGCTCTCCACAGCCCTGATTCGACTGCACCAATT	60
QY	1550	GCCCTGGGAATCCGGAATACATAGGGCTCTCCACAGCCCTGATTCGACTGCACCAATT	1609
Db	61	GCCCTGGGAATCCGGAATACATAGGGCTCTCCACAGCCCTGATTCGACTGCACCAATT	120
QY	1610	CTTGATTTGGGCGCTGTGCTGCTGCCTCATAGTATCTGCTTGGTCTTGGTGGGCGT	1659
Db	121	CTTGATTTGGGCGCTGTGCTGCTGCCTCATAGTATCTGCTTGGTCTTGGTGGGCGT	180
QY	1670	TCCAGGGGATGCTGCTTGGTTCAAGGACAAGACAGATGAAGGGTCTCACAGACACC	1729
Db	181	TCCAGGGGATGCTGCTTGGTTCAAGGACAAGACAGATGAAGGGTCTCACAGACACC	240
QY	1730	TGTTATCTCTTCTTTTCCACCTATCTCTCCACCCCTCCCTTTGCCCCACCAAG	1789
Db	241	TGTTATCTCTTCTTTTCCACCTATCTCTCCACCCCTCCCTTTGCCCCACCAAG	300
QY	1790	TTCCCATGTGCTGTACCTCCCTCCCTGTCTACATAGGACCTCTAGATAGTGTAGAGAG	1849
Db	301	TTCCCATGTGCTGTACCTCCCTCCCTGTCTACATAGGACCTCTAGATAGTGTAGAGAG	360
QY	1850	GAACATGTAGTGTAAATGAGTGTGGAATGGATT-GGCCCTCAGGCCAGGTGGTCTCAA	1908
Db	361	GAACATGTAGTGTAAATGAGTGTGGAATGGATTGGGCTCAGGCCAGGTGGTCTCAA	420
QY	1909	GGGACACGCTAACTGATCTTCCCTTTCAGAGACCCAGGAGTGGG---TTTCCCTCTT	1965
Db	421	GGGACACGCTAACTGATCTGCTCCCTTTCAGAGACCCAGGAGTGGGAGCTTCCCTCTT	480
QY	1966	CTCCAAGACTCAGGCTGTGGGCACTCTATAAGTAGTGTGCTCTCTCTGATAAC	2025
Db	481	CTCCAAGACTCAGGCTGTGGGCACTCTATAAGTAGTGTGCTCTCTCTGATAAC	540
QY	2026	AGATCAAAATTCCTCTCTCCCTCCACAGGTTTGGAAACAACTCTCCCTTCACTTGTG	2085
Db	541	AGATCAAAATTCCTCTCTCCCTCCACAGGTTTGGAAACAACTCTCCCTTCACTTGTG	600

Query Match	26.9%	Score 765	DB 13	Length 1201
Best Local Similarity	84.5%	Pred. No. 4.2e-192		
Matches 866	Conservative 35	Mismatches 118	Indels 9	Gaps 6

OY	38	TCCCGTCCCGGCGCGCGCGCGCGG-GGAAACAGCGGGG-CGGAGAGCGCTCCGGGAGC	95
Db	62	TCCGTTCCCGGCGCGCGCGCGCGGAGAACAGCGGGCGCCCGGGGGCGCCCGGGGGC	121
OY	96	GGGGGAGTTCGGGTTCCGGTTCTTTGTGCGGCTGCATCGAGGCTCCGGAGATGGC	155
Db	122	GGGGGGAGTTCCGGTTCCGGTTCTTTGTGCGGCTGCATCGAGGCTCCGGAGATGGC	181
OY	156	GCCCCGGCGGGTTTCCAGTCTGGCTCCAAACGGCGCGCGCGAGCTCAGAGGGGGC	215
Db	182	GCCCCGGCGGGTTTCCAGTCTGGCTCCAAACGGCGCGCGCGAGCTCAGAGGGGGC	241
OY	216	GGGCGGCTGCTGCTTTGGGAGCC-GGCGAATCCGGGGGCTCCTGTGCGAATGGGCG	274
Db	242	GGGCGGCTGCTGCTTTGGGAGCCCGGAGACTCCGGGGCTCTGTGTGCAATGGGCG	301
OY	275	GCTCCGGGTCAAGGGCTGTACCGCTCCCGATCCCGAGCGGCTTATCGAGCAAGT	334
Db	302	GCTCCGGGTCAAGGGCTGTACCGCTCCCGATCCCGAGCGGCTTATCGAGCAAGT	361
OY	335	ATGTGTCAAGGACGGCGGAATGACCTTCACTTCAAGGAGACCTTCAATGGAGCCCTGGGTATGG	394
Db	362	ATGTGTCAAGGACGGCGGAATGACCTTCAAGGAGACCTTCAATGGAGACCTTGGCTATGG	421
OY	395	GGGAACTCTTCAATCCGACTGGCTGGCCAGTCAAGGATGATCAATCCCGACAGAGA	454
Db	422	GGGAA-CTTCAATCCGACTGGCTGGCCAGTCAAGGATGATCAATCCCGACAGAGA	480
OY	455	CTGTGCTTCAAGATTCAGAGAGTCCAGACAGCGGGGTCCAAATTCGAATCCAAAT	514
Db	481	CTGTGCTTCAAGATTCAGAGAGTCCAGACAGCGGGGTCCAAATTCGAATCCAAAT	540
OY	515	GCAAGAGAGAGAGATGCGTCAAAATTTCTACTTCAAGATTTCTGTAACTGTACCA	574
Db	541	GCAAGAGAGAGAGATGCGTCAAAATTTCTACTTCAAGATTTCTGTAACTGTACCA	600
OY	575	GAATCCAGGCTATATGATCTCTTGCTTTGAAGAGAACTGACACAGACTATCATG	634
Db	601	GAATCCAGGCTATATGATCTCTTGCTTTGAAGAGAACTGACACAGACTATCATG	660
OY	635	AGGAAACGGTATATTCAGAGAGCTTGAAGCTCCATTAGCAAAAACGGAGCTG	694
Db	661	AGGAAACGGTATATTCAGAGAGCTTGAAGCTCCATTAGCAAAAACGGAGCTG	720
OY	695	CGAATTTTCAATTTCTAACAATTCAATCCGCTAAGTCAGATGCCGAGATGGGAGGG	754
Db	721	CGAATTTTCAATTTCTAACAATTCAATCCGCTAAGTCAGATGCCGAGATGGGAGGG	780
OY	755	ACGGTGGCTTCTGGGAGCTTGGGTAGAGAGCGGCTCTTGAAGATTCAGCTTGTCC	814
Db	781	ACGGTGGCTTCTGGGAGCTTGGGTAGAGAGCGGCTCTTGAAGATTCAGCTTGTCC	840
OY	815	AAATATGATGCACTAAACAAAGAGAGATTTCTCTTCTTTAATGCTCTTGGTAT	874
Db	841	AAATATGATGCACTAAACAAAGAGAGTCTC--TTCTTTTAAAGTCTTGTGT	897
OY	875	GAACTGACAAAGACCTGTATGGGCAGACAAACATCTGTAGATGGACAGAGCCGC	934
Db	898	GAACTGAC--CAARCTGTATGGGCAGAACCAATCTGTAGATGGACAGAGCCCACTAC	955
OY	935	ACTATCCAGAGACCGATGGCTTCCAGGTGAAGCGCGCAGAGATGTGAATGTACGGT	994
Db	956	AGGATCCCTGTATTAAGKACCGCGCGGAAAHKATGKVCGGGKGCATGTAGCGGAK	1011
OY	995	ACTGTCTTATGATGCTGATTAACAGGCTCCCAAGTTTAATATGACCCCGCTAGT	1051
Db	1016	VCAACCCCAAKGAGAGACCCCCCKRGSGMBCKGRACBMARCAAKKGRCAKARB	107
OY	1055	CGACTCTGTGGCATCATACCAACTC	1082
Db	1076	MMAKGCARGAAAAAARAAACAAACSC	1103

Db	301	TTGGATCAGTCCCGAAGAGACCTGCCCTCAGCAGATCCAGCAGGTCCAGCAGAGCGG	360
Qy	495	TCCAAAATCGAAAACCAAAATGCAAGAAAAAGAAAGATGGCTGCACAAAATTTACCTCAA	554
Db	361	TCCAAAATCGAAACCAAAATGCAAGAAAAAGAAAGATGGCTGCACAAAATTTACCTCAA	420
Qy	555	GGATTTCGTGAACCTGGTACCAAGATCCAGAGATCCAGAGCTATATGGATCTCTTTGGCTTTTGAAGGA	614
Db	421	GGATTTCGTGAACCTGGTACCAAGATCCAGAGATCCAGAGCTATATGGATCTCTTTGGCTTTTGAAGGA	480
Qy	615	AATGGACACAGACTATCATGAGGAAACGGCTAGATATCCAGAGCGCTTTGAAACGTCCTCA	674
Db	481	AATGGACACAGACTATCATGAGGAAACGGCTAGATATCCAGAGCGCTTTGAAACGTCCTCA	540
Qy	675	TTAAGCAAAAACGGAAGCTCGAATTTTTCATTTCTTAACACTTTTCAATCCGGCTAAGTCAG	734
Db	541	TCAAGCAAAAACGGAAGCTCGAATTTTTCATTTCTTAACACTTTTCAATCCGGCTAAGTCAG	600
Qy	735	ATGCCGAGGATGGGGAAGGACGGTGGCTTCTCGGAGCTTCGGGTAGAAAGACGGCTCC	794
Db	601	ATGCCGAGGATGGGGAAGGACGGTGGCTTCTCGGAGCTTCGGGTAGAAAGACGGCTCC	660
Qy	795	TGGAGGATTCAGCCTTGTCTCCAAATATATGACCATCAACAAAAGAGGAAAGTTCTTCTT	854
Db	661	TGGAGGATTCAGCCTTGTCTCCAAATATATGACCATCAACAAAAGAGGAAAGTTCTTCTT	720
Qy	855	TCTTTAAGTCTCTGGTCAATGAACTGGACAAAAGACCTGTATGGGCCAGACAACCAT--CT	912
Db	721	TTTTTTAAGTCTCTGGTCAATGAACTGGACAAAAGACCTGTATGGGCCAGACAACCATCTG	780
Qy	913	GGTAGAATGGCACAGGACCGCCAC	936
Db	781	GGTAGAATGGCACAGGACCGCC	804
RESULT 15			
BQ949766			
LOCUS			
DEFINITION	BQ949766	878 bp	linear
ACCESSION	BQ949766.1	GI:22365244	
VERSION	BQ949766.1	GI:22365244	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 878)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	Unpublished		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabs-remail.nih.gov		
	Tissue Procurement: DCTD/BTP/Gazdar		
	cDNA Library Preparation: Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LICM2554	row: p	column: 12
	High quality sequence	stomp: 693.	
FEATURES	Location/Qualifiers		
source	1..878		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6375035"		
	/tissue_type="large cell carcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_18"		
	/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:		



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 08:23:36 : Search time 4251 Seconds  
(without alignments/sec  
16590.966 Million cell updates/sec

Title: U66619

Perfect score: 1724  
Sequence: 1 GAATTCGCGCGAGCGGCC.....GTCAATCACTCGGAATTC 1724

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_inv:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hngo\_hum:\*  
40: em\_hngo\_mus:\*  
41: em\_hngo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1724	100.0	1724	9	U66619 Human SWI/S
2	1585.2	91.9	1698	9	BC002628 Homo sapi
3	1365.8	79.2	1413	12	BT007694 Synthetic
4	1198.6	69.5	1790	10	BC013122 Mus muscu
5	701.2	40.7	3086	10	BC026783 Mus muscu
6	690.6	40.1	2746	9	AF109733 Homo sapi
7	690.5	40.1	3335	9	BC009368 Homo sapi
8	685.2	39.7	3103	6	AX305314 Sequence
9	685.2	39.7	3103	10	MMU66620 Mus musculu
10	679	39.4	2747	5	BC049347 Danio rer
11	662.8	38.4	3101	10	MUSDI5X2L Mus musculu
12	655.4	38.0	2297	10	AB003504 Rattus no
13	649	37.6	2379	10	BC005732 Mus muscu
14	634.4	36.8	2041	9	HSU66618 Human SWI/S
15	630.6	36.6	2557	5	BC045009 Xenopus 1
16	572.6	33.2	2026	3	AY069383 Drosophi1
17	571	33.1	2018	3	AF071503 Drosophi1
18	523.8	30.3	2841	9	HSU66617 Human SWI/S
19	504.2	29.2	3344	2	AC014246 Drosophi1
20	504.2	29.2	171751	3	AC023697 Drosophi1
21	504.2	29.2	174157	3	AC023711 Drosophi1
22	504.2	29.2	332029	3	AE003491 Drosophi1
23	413.2	24.0	137956	2	AC010918 Drosophi1
24	410.4	23.8	450	6	AR240523 Sequence
25	405.6	23.5	1674	3	AK112677 Ciona int
26	276	16.0	1350	9	AK025917 Homo sapi
27	235.2	13.6	247	11	G30760 SW53165 Ex
28	234.8	13.6	44171	3	AF000265 Caenorhab
29	234.8	13.6	299719	2	AC006858 Caenorhab
30	216.6	12.6	294136	2	AC006901 Caenorhab
31	179	10.4	157075	2	AC005486 Homo sapi
32	174.2	10.1	214809	2	AC132744 Rattus no
33	174.2	10.1	220259	2	AC134763 Rattus no
34	174.2	10.1	263052	2	AC099360 Rattus no
35	170.4	9.9	35899	9	AC021097 Homo sapi
36	170.4	9.9	213045	2	AC093640 Homo sapi
37	167.8	9.7	27832	3	CEZK1128 Z47357 Caenorhabdi
38	167.8	9.7	239244	2	AC116136 Mus muscu
39	140.8	8.2	1216	17	AF113019 Homo sapi
40	137.8	8.0	310	6	AX098486 Sequence
41	136.8	7.9	160431	8	OSJN00176 A1662977 Oryza sat
42	133	7.7	1121	9	BC018853 Homo sapi
43	119.4	6.9	1898	8	AK119040 Arabidops
44	117.6	6.8	187859	2	BX323804 Danio rer
45	114.2	6.6	197095	2	BX322787 Danio rer

## ALIGNMENTS

RESULT 1  
HSU66619  
LOCUS HSU66619 1724 bp mRNA linear PRI 18-SEP-1996  
DEFINITION Human SWI/SNF complex 60 KDa subunit (BAF60c) mRNA, complete cds.  
ACCESSION U66619  
VERSION U66619.1 GI:1549246  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 1724)  
Wang, W., Cole, J., Xue, Y., Zhou, S., Khavari, P.A., Biggar, S.R.,  
Muehardt, C., Kalpana, G.V., Goff, S.P., Yaniv, M., Workman, J.L. and  
Crabtree, G.R.



QY 1501 TCCAGAGAGCCGTCAGTCTACTTCTACTGCAAGATCCAGAGCGAGGAGCTG 1560  
 |||||  
 DB 1501 TCCAGAGAGCCGTCAGTCTACTTCTACTGCAAGATCCAGAGCGAGGAGCTG 1560  
 |||||  
 QY 1561 GAGCAGTCGCTGTTGTGCGCAACCTAGAGCCCAAAACAGCAGCAGCGAAT 1620  
 |||||  
 DB 1561 GAGCAGTCGCTGTTGTGCGCAACCTAGAGCCCAAAACAGCAGCAGCGAAT 1620  
 |||||  
 QY 1621 TTCACCGCTGTCGCGGCGCCAGCATTTTGGCCCGGCTCCAGCTCCTCTGCGACC 1680  
 |||||  
 DB 1621 TTCACCGCTGTCGCGGCGCCAGCATTTTGGCCCGGCTCCAGCTCCTCTGCGACC 1680  
 |||||  
 QY 1681 TTGGGCTGTGGGGCTGATTAAAGCATTCATCTGGAATTC 1724  
 |||||  
 DB 1681 TTGGGCTGTGGGGCTGATTAAAGCATTCATCTGGAATTC 1724  
 |||||

RESULT 2  
 BC002628  
 LOCUS  
 DEFINITION BC002628 1698 bp mRNA linear PRI 12-JUL-2001  
 Homo sapiens, Similar to SWI/SNF related, matrix associated, actin  
 dependent regulator of chromatin, subfamily d, member 3, clone  
 MGC:3653 IMAGE:3608512, mRNA, complete cds.  
 BC002628  
 VERSION  
 KEYWORDS BC002628.1 GI:12803586  
 MGC.  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 1698)  
 TITLE  
 JOURNAL Strausberg, R.  
 Direct Submission  
 Submitted (05-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA

REMARK  
 COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Rubln Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mgc@hgrl.nih.gov](mailto:nisc.mgc@hgrl.nih.gov)  
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
 Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, R., Snyder, B., Stancirpop, S., Thomas, P.J.,  
 Tjongson, E.E., Touchman, J.W., Teurgeon, C., Vogt, J.L., Walker, M.A.,  
 Zhang, L.-H. and Green, E.D.

FEATURES  
 source  
 1..1698  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:3653 IMAGE:3608512"  
 /issue\_type="UTerms, endometrium adenocarcinoma"  
 /clone\_id="NIH\_MGC\_44"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 139..1551

CDS

/codon\_start=1  
 /product="Similar to SWI/SNF related, matrix associated,  
 actin dependent regulator of chromatin, subfamily d,  
 member 3"  
 /protein\_id="AA02628.1"  
 /db\_xref="GI:12803587"  
 /translation="MTPLQHPPTVVORPGMPGARMHQGAPMGPPYMGSPAVR  
 PGLAPAGMEPRKRAAPPGQSOAQSPVPTAPARSASAKRRKMDKILPQRIEL  
 VPESQAWMDILAPERKLDOTIMRKRDVIOALFKRPMQKRLRLVYSNTPNAPKPAE  
 DSDSISASWEELRVGSKLIDPSKOKRRSSFFKSLVLELDYLDYDPNDHLYBNHRTPT  
 TOEDYRQVKRPPGDLVSRCTLLMLDYQPPQFKLDPRLARLLGLHTOSRSALVQALWQ  
 YVKNTRLDSDHSEYINGDKYFOQIFDCPRLFSEIIPQRLTALLPDPPIYNHVISV  
 DPSDKTACYDIDVEVEEPLKQMSFSLSTANQOEISALDSKHIEFIESINDCKIO  
 RDPMLSPSRDQKGVVODLRLRSQSDLKMTMDVANPEERAEFEFHQWQSGEAVSRFY  
 YCKIQORROELQSLVAVNT"

BASE COUNT 386 a 535 c 480 g 297 t  
 ORIGIN

Query Match 91.9%; Score 1585.2; DB 9; Length 1698;  
 Best Local Similarity 99.3%; Pred. No. 3.8e-291;  
 Matches 1634; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

QY 79 CAGATCTCAAGATGCGCTTGTGTGTCGACAGCGCGTGGCTCCGGCCAGCAGCAG 138  
 |||||  
 DB 37 CAGATCTCAAGATGCGCTTGTGTGTCGACAGCGCGTGGCTCCGGCCAGCAGCAG 96  
 |||||  
 QY 139 GGGGCACTGATGATCACTTCCAGTGCAGACCCCTGCATCTATGACTCCAGTCTTCA 198  
 |||||  
 DB 97 GGGGCACTGATGATCACTTCCAGTGCAGACCCCTGCATCTATGACTCCAGTCTTCA 156  
 |||||  
 QY 199 CACCACCCACCGTGTGTCAGACGCCCGGAGATGCGGTGAGAGCCCGGATGCCACCA 257  
 |||||  
 DB 157 CACCACCCACCGTGTGTCAGACGCCCGGAGATGCGGTGAGAGCCCGGATGCCACCA 216  
 |||||  
 QY 258 GGGGCGCCATGAGGCGCCCGGCGCTCCCGTACATGAGGACCCCGCGTGGACCCGCG 317  
 |||||  
 DB 217 GGGGCGCCATGAGGCGCCCGGCGCTCCCGTACATGAGGACCCCGCGTGGACCCGCG 276  
 |||||  
 QY 318 CTGGCCCCCGGGG-ATGAGACCCCGCGCAAGCAGAGGCCCCCGCC-GGCAGAGC 375  
 |||||  
 DB 277 CTGGCCCCCGGGG-ATGAGACCCCGCGCAAGCAGAGGCCCCCGCC-GGCAGAGC 336  
 |||||  
 QY 376 CAGGACAGAGCCAGAGGCGCAGCCGAGCCAGCCCGCGCGGAGCGGAGTCCCAAG 435  
 |||||  
 DB 337 CAGGACAGAGCCAGAGGCGCAGCCGAGTCCAGCCGCGCGGAGCGGAGTCCCAAG 396  
 |||||  
 QY 436 AGAGAGAGATGGCTGACAAATCTCTCCCTCAAGAGATTTCGAGAGCTGCTCCGAGTCC 495  
 |||||  
 DB 397 AGAGAGAGATGGCTGACAAATCTCTCCCTCAAGAGATTTCGAGAGCTGCTCCGAGTCC 456  
 |||||  
 QY 496 CAGGCTTCAATGACCTTTGGCATTTAGAGGAACTGATCAAACTCATCTGCGGAAG 555  
 |||||  
 DB 457 CAGGCTTCAATGACCTTTGGCATTTAGAGGAACTGATCAAACTCATCTGCGGAAG 516  
 |||||  
 QY 556 GGGGTGACATCCAGAGAGCTCTGAAGAGGCCATTAAGCAAAAGCGGAGGTGGAGTCC 615  
 |||||  
 DB 517 GGGGTGACATCCAGAGAGCTCTGAAGAGGCCATTAAGCAAAAGCGGAGGTGGAGTCC 576  
 |||||  
 QY 616 TATATCTCAACACTTTTAACTTCGAGAGTCTGATGCTGAGATTCCGAGCGCAGATT 675  
 |||||  
 DB 577 TATATCTCAACACTTTTAACTTCGAGAGTCTGATGCTGAGATTCCGAGCGCAGATT 636  
 |||||  
 QY 676 GCTTCTGGAGCTACGGGTGAGAGGAACTCTTGATATCCAGCAAAACAGAGCGG 735  
 |||||  
 DB 637 GCTTCTGGAGCTACGGGTGAGAGGAACTCTTGATATCCAGCAAAACAGAGCGG 696  
 |||||  
 QY 736 AAGTCTCTTCTTCTTCAAGAGTTTGTGTCATCGAGCTGAGCAAAAGATCTTTAGGCTT 795  
 |||||  
 DB 697 AAGTCTCTTCTTCTTCAAGAGTTTGTGTCATCGAGCTGAGCAAAAGATCTTTAGGCTT 756  
 |||||  
 QY 796 GACAACCACTGTTGATGTCATCGACACCAAGCAGAGCAGAGCGGCTTCCAG 855  
 |||||  
 DB 757 GACAACCACTGTTGATGTCATCGACACCAAGCAGAGCAGAGCGGCTTCCAG 816  
 |||||

QY	856	GTGAAACGGCTGGGACCTGAGTGTGGGCTGCACGCTGCTCTCATCTGCTGACTACCAG	915
Db	817	GTGAAACGGCTGGGACCTGAGTGTGGGCTGCACGCTGCTCTCATCTGCTGACTACCAG	876
QY	916	CCTCCCAAGTTCAAACTGGATCCCGCTAGCCCGCTGCTGGGGCTGCACACACAGAGC	975
Db	877	CCTCCCAAGTTCAAACTGGATCCCGCTAGCCCGCTGCTGGGGCTGCACACACAGAGC	936
QY	976	CGCTCAGCCATTGTCACGCGCTGGGAGATGTGAAGACCAACAGAGCTCAGGACTCC	1035
Db	937	CGCTCAGCCATTGTCACGCGCTGGGAGATGTGAAGACCAACAGAGCTCAGGACTCC	996
QY	1036	CATGACAAAGGAATACATCAATGGGACCAAGTATTTCCAGCAGATTTTTCATTTGCTCCCGG	1095
Db	997	CATGACAAAGGAATACATCAATGGGACCAAGTATTTCCAGCAGATTTTTCATTTGCTCCCGG	1056
QY	1096	CTGAAGTTTCTGAGATTTCCCGCTGCTCAGCGCTGCTATTTGCCCTGACCCAAATT	1155
Db	1057	CTGAAGTTTCTGAGATTTCCCGCTGCTCAGCGCTGCTATTTGCCCTGACCCAAATT	1116
QY	1156	GTCAACCAACATGTCATCAGGTGACCCCTTCCAGACCAAGAGACACAGCGTGTATGAC	1215
Db	1117	GTCAACCAACATGTCATCAGGTGACCCCTTCCAGACCAAGAGACACAGCGTGTATGAC	1176
QY	1216	ATTGACGTGGAGTGGAGAGCATTAAAGGGGACAGATGAGCAGCTTCTCTATCCACG	1275
Db	1177	ATTGACGTGGAGTGGAGAGCATTAAAGGGGACAGATGAGCAGCTTCTCTATCCACG	1236
QY	1276	GCCAAACAGCAGGAGATCAGTCTCTGACAGTAAGATCCATGAGACGATTTGAGTCCATA	1335
Db	1237	GCCAAACAGCAGGAGATCAGTCTCTGACAGTAAGATCCATGAGACGATTTGAGTCCATA	1296
QY	1336	AACCAAGCTCAGATCAGAGGACCTTCAAGCTTCTCAGAGACCCCAAGGCTAT	1395
Db	1297	AACCAAGCTCAGATCAGAGGACCTTCAAGCTTCTCAGAGACCCCAAGGCTAT	1356
QY	1396	GTCCAAAGACCTCTCGCTCCAGAGCGGACCTCAAGGTCATGACAGATGTAGCCGGC	1455
Db	1357	GTCCAAAGACCTCTCGCTCCAGAGCGGACCTCAAGGTCATGACAGATGTAGCCGGC	1416
QY	1456	AACCTCAGAGAGCGCGGCTGAGTTTACCAACAGCCCTGTGTCAGAGAGCCGCTC	1515
Db	1417	AACCTCAGAGAGCGCGGCTGAGTTTACCAACAGCCCTGTGTCAGAGAGCCGCTC	1476
QY	1516	AGTCCTACTTCTACTGCAAGATCAGCAGCGCAGGAGCTGAGCAGTGCCTGGTT	1575
Db	1477	AGTCCTACTTCTACTGCAAGATCAGCAGCGCAGGAGCTGAGCAGTGCCTGGTT	1536
QY	1576	GTGCGCAACACCTAGGAGCCCAAAACAGCAGCAGCGAACTTTTCAGCCGTGTCGCG	1635
Db	1537	GTGCGCAACACCTAGGAGCCCAAAATAGCAGCAGCGAACTTTTCAGCCGTGTCGCG	1596
QY	1636	GGCCCCAGATTTTCCCGGCTCCAGC-TCACCTCTCTGCACTTTGGGTGTGGGC	1694
Db	1597	GGCCCCAGATTTTCCCGGCTCCAGC-TCACCTCTCTGCACTTTGGGTGTGGGC	1656
QY	1695	TGGATTAAAGTCATCTCTGGAA	1720
Db	1657	TGGATTAAAGTCATCTCTGACAA	1682

RESULT 3  
BT007694  
LOCUS  
DEFINITION  
BT007694 1413 bp mRNA linear SYN 13-MAY-2003  
Synthetic construct Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3 mRNA, partial cds.  
ACCESSION  
BT007694  
VERSION  
BT007694.1  
KEYWORDS  
FLI\_CDNA.  
SOURCE  
synthetic construct  
ORGANISM  
synthetic construct

artificial sequences. 1 (bases 1 to 1413) Kalinine,N., Chen,X., Rolfes,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A. Cloning of human full-length CDSs in BD Creator(TM) System Donor vector 2 (bases 1 to 1413) Kalinine,N., Chen,X., Rolfes,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A. Direct Submission Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: http://bioinfo.clontech.com/orfclones. Location/Qualifiers 1..1413 /organism="synthetic construct" /mol_type="mRNA" /db_xref="taxon:32630" /clone="GH00091L1.0" /clone_lib="BD Creator(TM) CDS Library derived from MGC collection" /lab_hosts="DH5alpha T1 resistant" /note="vector: pDNR-Dual" 1..>1413 /note="Mutations: 12:OK;1412:Stop->Leu" /transl_table=1 /product="Homo sapiens SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3" /protein_id="AAP36362.1" /db_xref="GI:30584227" /translation="MTPLQLHPVTWVPGMPSGARMPHQAGPMGPPGSPYMGSPAVR PGLAPAGMEAPKRAEAPPQSOAQSGQVPVTPAPRSARAKRMADKILPQRIREL VPESQAYMDLLAFERKLDQTIMRKVRDIQEALEKPMQKRLRLIYISNTPNAPKDAE TSDGSIAWELRVGKLLDDPSKQRRKFSFSLVLELDKLYGPDNHLVWHRTPT QVTDGQVQRKPGDLSVRCITLLMLDYQPPQKLDPLRLALHLSRSLVQALWQ YVKTNRQLQSDHKEYINGDKYFQIIPDCPRKFSEIPQRLTLLPDPPIVINHVIS DPSDKTACVDIDVEVEPLKGMSSFLSTANQOBSIALDSKHITIEINOLKIQ RDFMLSPSRQPKGVVQDLRSQSRDLKVMITDVAGNPEERAEFYHQPSQEAHSRYF YCKTQRQRLQLLEQLSLVRNTLI" 320 a 446 c 404 g 243 t	BASE COUNT ORIGIN Query Match 79.2%; Score 1365.8; DB 12; Length 1413; Best Local Similarity 99.3%; Pred. No. 1.9e-249; Matches 1403; Conservative 0; Mismatches 7; Indels 3; Gaps 3; QY 181 ATGACTCCAGTCTTTCAGCACCCACCCACCCGCGGCTCCCGGATGCCGCTCGGA 240 Db 1 ATGACTCCAGGCTTTCAGCACCCACCCACCCGCGGCTCCCGGATGCCGCTCGGA 60 QY 241 GCCCGGATGCCACCA-GGGGCGCCCATGGGCCCCCGGGCTCCCCCTACATGGGCAGC 299 Db 61 GCCCGGATGCCACCAAGGGGGCGCCCATGGGCCCCCGGGCTCCCCCTACATGGGCAGC 120 QY 300 CCGCGCGTGGACCCCGGCTGGCCCCCGCGGG-ATGGAGCCCGCCGCAAGGACGACGG 358 Db 121 CCGCGCGTGGACCCCGGCTGGCCCCCGCGGGCATGGAGCCCGCCGCAAGGACGACGG 180 QY 359 CCGCGCGCC-GGCAGAGCCAGGCACAGAGCAGGGCCAGCGCCGAGCCACCGCCCCCGG 417
--	---

```

Db      181  |CCCCCGGAGAGAGCCAGAGGCAAGAGGCGGAGCGGAGTCCACCGCCCCGGG|
Qy      418  |CGGAGCGGAGTCCCAAGAGAGAGAGATGGCTGACAAATAATCTCCCTCAAGATTGG|
Db      241  |CGAGCGGAGTCCCAAGAGAGAGAGATGGCTGACAAATAATCTCCCTCAAGATTGG|
Qy      478  |GAGCTGGTCCCGAGTCCGAGGCTTTCATGAGACCTCTTGGCATTTGAGAGAACTGAT|
Db      301  |GAGCTGGTCCCGAGTCCGAGGCTTTCATGAGACCTCTTGGCATTTGAGAGAACTGAT|
Qy      538  |CAACCATATGTCGGAAGGGGGTGAATCCAGAGAGCTCTGAAGAAGCCATGAAGCA|
Db      361  |CAACCATATGTCGGAAGGGGGTGAATCCAGAGAGCTCTGAAGAAGCCATGAAGCA|
Qy      598  |AAGCGGAGCTGCGACTTATATCTTCAACATTTTAACTTCCGCAAGTCTGATGCTG|
Db      421  |AAGCGGAGCTGCGACTTATATCTTCAACATTTTAACTTCCGCAAGTCTGATGCTG|
Qy      658  |GATTCGAGAGGAGCATTCGCTCTGAGAGCTACGGGTGAGAGGGAAGCTCTGATGAT|
Db      481  |GATTCGAGAGGAGCATTCGCTCTGAGAGCTACGGGTGAGAGGGAAGCTCTGATGAT|
Qy      718  |CCGAGCAAAAGAGAGGAGGAGTCTCTCTTCTTCAAGAGTTTGTCAATGAGCTGAG|
Db      541  |CCGAGCAAAAGAGAGGAGGAGTCTCTCTTCTTCAAGAGTTTGTCAATGAGCTGAG|
Qy      778  |AAAGATCTTATATGCGCTTCAACACCTCTGAGTGGATGAGTGGACCCAGACCCG|
Db      601  |AAAGATCTTATATGCGCTTCAACACCTCTGAGTGGATGAGTGGACCCAGACCCG|
Qy      838  |GAGAGGAGAGGCTTCCAGGTGAAGGGGCGGAGCTGAGTGGCTGACGCTGCTC|
Db      661  |GAGAGGAGAGGCTTCCAGGTGAAGGGGCGGAGCTGAGTGGCTGACGCTGCTC|
Qy      898  |CTCATGCTGAGTACCAAGGCTCCCACTTCAAACTGATCCCGCTTACCGGCTGCTG|
Db      721  |CTCATGCTGAGTACCAAGGCTCCCACTTCAAACTGATCCCGCTTACCGGCTGCTG|
Qy      958  |GGGCTGCAACAGAGCGGCTGAGCATTTGTCAGAGCCCTGAGAGTGTGAAGACC|
Db      781  |GGGCTGCAACAGAGCGGCTGAGCATTTGTCAGAGCCCTGAGAGTGTGAAGACC|
Qy      1018 |AACAGCTGAGAGCTCCCATATGACAAAGAAATCATATGGGAGCAAGATTTCCAGC|
Db      841  |AACAGCTGAGAGCTCCCATATGACAAAGAAATCATATGGGAGCAAGATTTCCAGC|
Qy      1078 |ATTTTGAATTTGTCGCGGCTGAAGTTTCTGAGATTCGCGAGGCTTCAAGGCTGCTA|
Db      901  |ATTTTGAATTTGTCGCGGCTGAAGTTTCTGAGATTCGCGAGGCTTCAAGGCTGCTA|
Qy      1138 |TTGCCCCCTGAGCCCAATTTGTCATCAACATGTCATCAGGTGAGACCTTCAAGC|
Db      961  |TTGCCCCCTGAGCCCAATTTGTCATCAACATGTCATCAGGTGAGACCTTCAAGC|
Qy      1198 |AAGACAGCTGCTATATACATTTGACGTGAGGTGAGAGACCATTAAGGGGAGATGAG|
Db      1021 |AAGACAGCTGCTATATACATTTGACGTGAGGTGAGAGACCATTAAGGGGAGATGAG|
Qy      1258 |AGCTTCTCTTATCCAGGCGCAACAGAGAGATGAGTCTCTGAGACATTAAGGCTG|
Db      1081 |AGCTTCTCTTATCCAGGCGCAACAGAGAGATGAGTCTCTGAGACATTAAGGCTG|
Qy      1318 |GAGAGATTTGAGTCCATAAAGAGCTCAAGATTCAGAGGAGCTTCAAGCTTCTCC|
Db      1141 |GAGAGATTTGAGTCCATAAAGAGCTCAAGATTCAGAGGAGCTTCAAGCTTCTCC|
Qy      1378 |AGAGAGCCCAAGGCTATATGTCAGAGCTGCTCCGCTCCCAAGGCGGAGCTCAAG|
Db      1201 |AGAGAGCCCAAGGCTATATGTCAGAGCTGCTCCGCTCCCAAGGCGGAGCTCAAG|
Qy      1438 |ATGACAGATGTAAGCGGCAACCTGAAGAGAGCGCGGAGTGTACACACCGAGCC|

```

```

Db      1261 |ATGACAGATGTAGCCGCAACCTTGAAGAGAGCGCGGAGTGTACTTACCAACAGCCC|
Qy      1498 |TGGTCCGAGAGAGCGGCTGAGTCTTCTATCTGCAAGATTCACAGCGGAGGAGAG|
Db      1321 |TGGTCCGAGAGAGCGGCTGAGTCTTCTATCTGCAAGATTCACAGCGGAGGAGAG|
Qy      1558 |CTGAGAGAGTCTGCTGTTGTGGCAACCTTAG|
Db      1381 |CTGAGAGAGTCTGCTGTTGTGGCAACCTTAG|

RESULT 4
BC013122
LOCUS
DEFINITION
Mus musculus SWI/SNF related, matrix associated, actin dependent
regulator of chromatin, subfamily d, member 3, mRNA (CDNA clone
MGC:8018 IMAGE:3586276), complete cds.
ACCESSION
BC013122
VERSION
BC013122.1 GI:15341882
KEYWORDS
MGC.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 1790)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L.H., Shemen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Scheffer, C.F., Bhac, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Matushina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schreib, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Rana, S.S., Iqbal, N.A., Peters, G.J.,
Abramson, R.D., Wulliam, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahy, J., Hellmuth, E., Ketterman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shcherbakov, Y.,
Boutard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Gilwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalley, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL
MEDLINE
PUBMED
22388257
12477932
2 (bases 1 to 1790)
REFERENCE
Strausberg, R.
Direct Submission
Submitted (27-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgepds-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILN)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
Gunsberg, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louie, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILN at: http://image.llnl.gov
Series: IRAK Plate: 10 Row: 1 Column: 23

```

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 133853379.

FEATURES	source	Location/Qualifiers
gene		1..1790 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="MGC:8018 IMAGE:3586276" /tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy." /clone_lib="NCI CGAP Mam1" /lab_host="DH10B" /note="Vector: pCMV-SPORT6"
CDS		1..1790 /gene="Smarcd3" /db_xref="LocusID:6993" /db_xref="MGI:1914243" 488..1633 /codon_start=1 /product="Smarcd3 protein" /protein_id="AAH13122.1" /db_xref="GI:15341883" /db_xref="LocusID:6993" /translation="MADKILPQRIRELVPSQAYMDLLAPERKLDQTIMKRKRVDIQSAFLKRPKMKQKRLRYISINTFAPKPAEDSDGSASWELRVGKLLDDPSQKRRKFSFQFKSLVIDLKDLYGPDNHLVWEHRTPTTQETDQFVKRPDGLSVRCVTLMLMDYQPPQFKLIDPLRLGLLGHQTQSRSAIYOALWOYVNTNRNQDSDHKEYINGDKYFOOIIPDCPLKFKSEIPORLTALLPPDPIVINHVISVDPDQKKTACVDIDVEEPLKQGMSSFLLS TAQOEISALDSKTHETIESINQLKIQDPMFLSPSRDPKGYVQDLLRSQSRDLKVMYTDVAGNPEERAEFYHQWSEAEVSYFYFKIQRRQLEQSLVVRNT"
BASE COUNT	447 a	522 c 516 g 305 t
ORIGIN		
Query Match	69.5%	Score 1198.6; DB 10; Length 1790;
Best Local Similarity	92.1%	Pred No.1.le-217;
Matches 1297; Conservative	0;	Mismatches 109; Indels 3; Gaps 3;
Qy	219	GGCGCCCGGGATGCGGTGTGGAGCCCGGATGCCCCACCA-GGGCGCGCCCATGGCCCCCCCC 277
Db	259	GGCGCCCGGGATGCGGTGTGGAGCCCGGAATGTCCTCCACAGGGGCGCCCATGGCCCCCCCC 318
Qy	278	GGGCTCCCCGTACATGGGCAGCCCCCGTCGTGCGACCCGGCTGCGCCCCCGGG-ATGGA 336
Db	319	GGGCTCCCCGTACATGGGCAGCCCCCGGTACGACCCGGCTGCGCCCCCGGGCATGGA 378
Qy	337	GCCCGCCGCAAGGAGCAGCGCCCCCGCCCC-GGCAGAGCCAGGACACAGAGCCAGGGCCA 395
Db	379	GCCCGCCGCAAGGAGCAGCGCCCCCGCCCGCGGCGAGCAGGACAGGGCCAGGGCCA 438
Qy	396	GCGGAGCCCAACCGCCCCCGCGCGGAGCGCGAGTGCACAGAGGAGGAAGATGGCTGACAA 455
Db	439	GCCCGTCCCAACCGCCCCAGCGCGAGCGCGAGTGCACCAAGAGGAGGAAGATGGCTGCAA 498
Qy	456	AATCTCTCCCTCAAGAGATTGGGAGCTGGTCCCGAGTCCCGAGCTTACATGGACCTCTT 515
Db	499	AATCTCTCCCTCAAGAGATTGGGAGCTGGTACCCGAGTCCCGAGCTTACATGGACCTCTT 558
Qy	516	GGCATTGTAGAGGAAAATGGATCAAAACCATCATCGGAAGGGGGTGGACATCCAGGAGGC 575
Db	559	AGCATTGTAGAGGAAAATGGATCAAAACCATCATCGGAAGCGGTGGACATCCAGGAGGC 618
Qy	576	TCTGAAGAGCCCATGAAGCAAGCGGAGCTCGGACTCTATATCTTCCAAACATTTTAA 635
Db	619	CCTGAAGAGCCCATGAAGCAAGCAAGCGGAGCTCGGCTTTATATCTTCCAACTACTTTAA 678
Qy	636	CCCTGCCAAGTCTGATCGTGAAGATTCCGACGGCAGCATTCCTCTCTGGGAGTACCGGT 695
Db	679	CCCTGCCAAGCCTGATCGGAAGACTCTGATGGCAGCATTCCTCTCTGGGAGTCCGGGT 738
Qy	696	GGAGGGGAAGCTCTCTGGATGATCCAGCAAAAGAGCGGAAGTCTCTCTTTCTTTCAA 755
Db	739	GGAGGGGAAGCTCTCTGGATGATCTAGTAAGCAGAGGAAGTTTCTTCTTCTTTCAA 798

REFERENCE 1 (bases 1 to 3086)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LIML)  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
 Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louieged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>  
 Series: IRAX Plate: 41 Row: h Column: 9  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13937376.  
 Location/Qualifiers  
 1..3086  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4481097"  
 /tissue\_type="Eye, retina, mouse strain C57BL/6"  
 /clone\_id="N1H\_MGC\_94"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 <1..1495  
 /codon\_start=2  
 /product="Unknown (protein for IMAGE:4481097)"  
 /protein\_id="AAH26783.2"  
 /db\_xref="GI:27693623"  
 /translation="SGAGVAAALGPGTGPVPMGAPGGLYRSPWGAAYPRPGMLGSRMTPOGSPGPGVGNASVPRGLAOSGMDOSRPRPPOIIOVOOAOVONRNHNKKKKMADKILPORIRELVESSQAYMDLPERKLDOTIRKSLDIDELAKRIKOKKRLIFISNTNPAKSDAEDGTVASVELRVEGLSDALSYDADKQKRTSSFFKSLVLELDKODYGPDMHVEWMTATTDGDFQVKGEDVNVACTVLMIDYQPOFKLIDRLARLGIHTQTRPVIIOALMOYIKTHLQDPHREFFVLCDKYLOQIFESQRMKFSLEIQRHLALMPEPIIINHVISVDNDOKTACVYIDVEDDTLKTOMNSFLSTASOEIATLNDKIHETIETINOLKTQREPMSPARDPGFIDMDILOSCKRLKMTDVTGNEBEERRAEFYQPMAOEAVCRFYFSKVQROEALGIRNT"

BASE COUNT 680 a 911 c 798 g 697 t  
 ORIGIN

Query Match 40.7%; Score 701.2; DB 10; Length 3086;  
 Best Local Similarity 71.3%; Pred. No. 3.8e-123;  
 Matches 1000; Conservative 0; Mismatches 358; Indels 44; Gaps 4;

QY 217 CAGCGCCCGGAGTGCCTCTGAGGCCCGGATGCCACACAGGGGCGCCCATGGGCC 276  
 DB 122 CCGAGACCAAGTATGTGTCAGGTAGCGAATGACACCTCAGGGAACCTTCATGGAACT 181  
 QY 277 CGGGCTCCCGGTCATGGGAGCGCCGCGGTGACCCGCGCTGGCCCCCGC-GGGATGG 335  
 DB 182 C-----CTGGCTATGGGGGAAACCTTCAGTCCGACCTGTCTGGCCCAAGTATGG 236  
 QY 336 AGCCGCGCCGCAAGAGCAGCGCGCCCGCGGAGAGCAGGACAGCAGAGCGAGCGCA 395  
 DB 237 ACCAGTCCCGCAAGAGACTGCACCTC-----AACAGATCCAGAGGT 279  
 QY 396 GCCGAGCCACCGCCCGCGCGAGCCGACGCTGCGCAAGAGAGAGATGGCTGCA 455  
 DB 280 CCAGCAGCAGCGGCTCAAAATCGAAATCACAATGCAAAAGAAAGATGGCTGCA 339

QY 456 AATCTCCCTCAAGAGATTGGGAGCTGTGCCAGTCCAGGCTTACATGACCTCTT 515  
 DB 340 AATCTTACTCAAGAGATTGGGAACTGTGCCAATATACAGGCTTACATGATCTCTT 399  
 QY 516 GGCATTGAGAGAAATCGATCAAAACATCATGCGGAAGGGGGTGGACATCCAGAGGC 575  
 DB 400 GGCATTGAGAGAAATCGACACAGACTATATGAGAAAGCGGTAGATATCCAGAGGC 459  
 QY 576 TGTGAAGAGCCCATGAAGCAAAAGCGAGTGGACCTCTATCTCTCAACCTTTTA 635  
 DB 460 CTGGAAGCGTCCCATCAAGCAAAAGCGAGTGGAAATTTTCATTTCAACAGCTTCA 519  
 QY 636 CCTCGGAAGTGTGATGCTGAGATTCGACCGCGACATTCGCTCGGAGCTACGGT 695  
 DB 520 TCCGCTAAGTGGACGCGGAGATGGGGAAGGACGCTTCTCTGGAGCTCCGGT 579  
 QY 696 GGAGGGGAAGCTCTGG-----ATGATCCAGCAACAGAGCG 734  
 DB 580 AGAAGCGCGGCTCTGGAGAGCGCGGCTGTCCAAATATGACCGCACCAAGCAAAAG 639  
 QY 735 GAAGTTCTCTTTTCTTCTCAAGATTTGTCTATGAGCTGACCAAAAGATCTTTATGGCC 794  
 DB 640 AAAGTTCTCTCTTTTAACTCTTGTATGCAACTGACCAAGACTGATAGCC 699  
 QY 795 TGACAACACCTGTTGATGCGATCGGACACCCAGACCCAGAGAGAGCGGCTTCA 854  
 DB 700 AGACAACATCTGTGATGATGACACAGACCGCCACTACCGAGAGACCGATGCTTCA 759  
 QY 855 GGTGAAGCGGCTGGGAGCTGAGTGTGGCTGACAGCTGCTCTCATGCTGGACTACA 914  
 DB 760 GGTGAAGCGGCTGGAGAGATGTAATGTACGGTGTACTGTCTGTGATGCTGGACTACA 819  
 QY 915 GCCTCCCAAGTTCAACTGATATCCCGCTACCGCGCTGCTGGGCTGCAACACAGG 974  
 DB 820 GCCCCCCAGTTTAATTAGACCTCTGCTGCTGGCTCTTGGGCTATCAATACAGAC 879  
 QY 975 CCGCTCAGCATTTGCGAGGCGCTGGGAGATGTGAAGACCAACAGGCTCAGAGATC 1034  
 DB 880 AGCTCAGGATCATCAAGACCTGGGAGTATTTAAACACCAAGCTCAGAGATC 939  
 QY 1035 CCATGACAAGATCATCAATGAGGAGATTTCCAGAGATTTTGTATGTTGCCCG 1094  
 DB 940 TCAGAGGAGAGATTTGTTCTCTGTGACAGTACTCCAGAGATCTTGAATCTCAGG 999  
 QY 1095 GCTGAAGTTTGTGATTTCCAGCGCTCAACAGCTCTGATTTGCCCTGACCCAT 1154  
 DB 1000 GATGAAGTTTCAAGATCCCTCAGCGGCTCAGCGCTTGTATGCAACAGAGCCAT 1059  
 QY 1155 TGTATCAACATGATCAAGGTGACCTTCAACAGAGAGAGAGAGAGCGGTATGA 1214  
 DB 1060 CATCATCAATCATGATCATGATGAGCCCAATATACAGAAAGACCGGCTGATGA 1119  
 QY 1215 CATTGACGTGAGTGAAGGAGCCATTAAGGGGAGATGAGCAGCTTCTCTATCCAC 1274  
 DB 1120 CATTACGTGAGTGAAGTGAACATCTGTAAGACCCAGATGAATCTTCTTGTGTCAC 1179  
 QY 1275 GGCACAAGCAGAGATCACTCTCTGAGACAGTAAGATTCATGAGAGATTTAGTCAT 1334  
 DB 1180 TGCCAGCAGCAGAGATGCGACACTGACACAAAGATTCATGAGAGATTCAT 1239  
 QY 1335 AAACAGCTCAAGATTCAGAGAGATTCATGTAAGCTTCTTCAAGACCCCAAGGCTA 1394  
 DB 1240 CAACAGCTGAAGACCCAGAGATTCATGTTGAGCTTGTGCGAGACCTCAGAGGTTT 1299  
 QY 1395 TGTCCAAAGCTGTCTCCCTCCAGAGCGGAGCTCAAGGTATGAGATGATGAGCGG 1454  
 DB 1300 CATCATGATTTGGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1359  
 QY 1455 CAACCTGAAGAGAGCGCGGCTGAGTTCTTACACAGAGCTGTGCTCAGAGAGCGCT 1514  
 DB 1360 TAAACCGAAGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1419



```

Qy 1515 CAGTCGCTACTTCTACTCAAGATCCAGCAGCGCAGGAGGAGCTGGAGCGATCGCTCGT 1574
Db 1420 GTGCCGATATTTCTACTCAAGGTGCAGCAGGCGCGCAAGAGTTAGAGCAAGCCTCGG 1479
Qy 1575 TGTGCGCAACACTAGGAGCCC 1596
Db 1480 AATCCGAACACATAGGCGCTC 1501

RESULT 6
AF109733 2746 bp mRNA linear PRI 06-APR-1999
LOCUS Homo sapiens SWI/SNF-related, matrix-associated, actin-dependent
DEFINITION regulator of chromatin D1 (SMARCD1) mRNA, complete cds.
ACCESSION AF109733
VERSION AF109733.1 GI:4566529
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2746)
AUTHORS Wang, W., Xue, Y., Zhou, S., Kuo, A., Cairns, B. R. and Crabtree, G. R.
TITLE Diversity and specialization of mammalian SWI/SNF complexes
JOURNAL Genes Dev. 10 (17), 2117-2130 (1996)
MEDLINE 96397413
PUBMED 8804307
REFERENCE 2 (bases 1 to 2746)
AUTHORS Ring, H. Z., Vameghi-Meyers, V., Wang, W., Crabtree, G. R. and Francke, U.
TITLE Five SWI/SNF-related, matrix-associated, actin-dependent regulator
of chromatin (SMAR) genes are dispersed in the human genome
JOURNAL Genomics 51 (1), 140-143 (1998)
MEDLINE 98360103
PUBMED 9693044
REFERENCE 3 (bases 1 to 2746)
AUTHORS Ring, H. Z. and Francke, U.
TITLE Direct Submission
JOURNAL Submitted (25-NOV-1998) Genetics and HHMI, Stanford University
School of Medicine, B205 Beckman Center, Stanford, CA 94305, USA
FEATURES
source
1..2746
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q13-q14"
/tissue_type="peripheral blood"
/note="from two individuals"
1..2746
/gene="SMARCD1"
145..1506
/gene="SMARCD1"
/note="BAF60a; similar to sequences presented in GenBank
Accession Numbers U66617, U66618, U66619 and U66620"
/codon_start=1
/product="SWI/SNF-related, matrix-associated,
actin-dependent regulator of chromatin D1"
/protein_id="AAD23390.1"
/db_xref="GI:4566530"
translation="MLPGSRMTPOGSPGPPYGNPSVRPGLAQSGMDQSRKRPAPQ
QIQVQQAQVQNRHNKAKKMKADILPQIRRELVPESQAYMDLLAFERKLDQTIMRK
LIDQALKRPIQOKRKRIPIFNTFPAKDAEDGEGTVASWELRVGRLLDSALS
KYDATKOKRKFSSFFKSLVIELDKLVGPDNLVWHRHTATQETDGFQVKRPGDVNV
RCTVLLMDYDPPQKLPRLARLLGIHTQRPVILQALWQYIKTKLQDPERFVI
CDKYLQIFESQRMKFSPIQRLHALLMPEPIINHVISVDPNDQKTKATYDIDVEV
DDTLKQNSFLLSTASQEIATDNKIHTETINQLTKRFLMSLPDQPFIND
WLQSQCDRLKTMVDVGNPEBERAEFFQFQWQAQVAVCRYPYFQVQRQRLQALGI
RNT"
misc_difference 1228..1350
/gene="SMARCD1"
/note="difference between this sequence and Homo sapiens
sequence presented in GenBank Accession Number U66617"
/replace=""

```

```

BASE COUNT 628 a 777 c 686 g 655 t
ORIGIN
Query Match 40.1%; Score 690.6; DB 9; Length 2746;
Best Local Similarity 70.7%; Pred. No. 3.9e-121;
Matches 996; Conservative 0; Mismatches 369; Indels 44; Gaps 4;
Qy 217 CAGCGCCCGGGATGCGCTCTGGAGCCCGGATGCCACCAAGGGGCGCCCATGGCCCCC 276
Db 133 CCGAGACCAAGTATGTTGCCAGGCAGCGAATACACTCAGGACCTTCCATGGGACCC 192
Qy 277 CGGCTCTCCCGTACATGGGCGACCCCGCGTGCACCGGCTGGCCCCCGC-GGGATGG 335
Db 193 C-----CTGGCTATGGGGGAACTTTCAGTCCAGCTGGCGCTGGCCCGCAGTCAGGGATG 247
Qy 336 AGCCCGCCGCAAGCGAGCGCGCCCGCCCGCCGCGCAGAGCCAGGACACAGACCCAGGGCCA 395
Db 248 ATCAGTCCCGCAAGACCTGCCCTCAG-----CAGATCCAGCAGGT 290
Qy 396 GCCGAGGCCACCCCGCCCGCGAGCGCGAGTCCCAAGAGGAGGAGATGGCTGACAA 455
Db 291 CCAGCAGCAGCGCTCCAAAATCGAAACCAATGCAAAAGAAAAGAAAGATGGCTGACAA 350
Qy 456 AATCCTCCTCAAGGATTCGGGAGCTGGTCCCGAGTCCCGAGTCCAGGCTTACATGGACCTTT 515
Db 351 AATTCTACTCAAGAGATTCTGTAACGTGTACCAAGAAATCCAGGCCTATATGATCTCTT 410
Qy 516 GGCAATTTGAGAGGAAACTGGATCAAAACCATCATCGGAAGGGGTGGACATCCAGGAGGC 575
Db 411 GGCTTTTGAAGGAAACTGGACAGATATCATGAGGAAACGGCTAGATATCCAGAGGC 470
Qy 576 TCTCAAGAGGCCCATGAAGCAAAAGCGAAGCTCGACTCTATATCTCTCAACACTTTTAA 635
Db 471 CTTGAACGTCCTCATCAAGCAAAAACGGAAGCTCGGAATTTTCATTTCTAACACTTTCAA 530
Qy 636 CCCTGCGAAGTCTCATGCTGAGGATTCGACGGCAGCATTCGCTCTCGGGAGCTACGGGT 695
Db 531 TCCGCTCAAGTCAGATGCGAGGATGGGAAGGACGGTGGCTTCTCTGGGAGCTTCGGGT 590
Qy 696 GGAGGGAGCTCCTGG-----ATGATCCAGCAACAGAAAGCG 734
Db 591 AGAAGAGCGGCTCCTGGAGGATTCAGCTTGTCCAATATATGTCCTAACAAAGAG 650
Qy 735 GAAGTTCTCTCTTTCTTTCTTTCAAGAGTTTGGTTCATCGAGCTGCAACAAAGATCTTTATGGCCC 794
Db 651 GAAGTTCTCTCTTTCTTTCTTTTAACTCTTGGTGAATGAACTGCAACAAAGACCTGTATGGGCC 710
Qy 795 TGACAAACACCTCTGTTGAGTGGCATTCGGACACCCACGACCCAGGAGCGGCTTCCA 854
Db 711 AGACAAACCATCTGGTAGAATGGCACAGGACCGCCACTACCCAGGAGACCGATGGCTTTCA 770
Qy 855 GGTGAACCGGCTGGGACCTGAGTGTGGCTGCACGCTGCTCTCATGCTGGACTACCA 914
Db 771 GGTGAACCGGCGGGAGACGTAATGTAACGCTGTAAGTGTCTCTGATGCTGGATTTACCA 830
Qy 915 GCCTCCCGAGTTCAAACTGGATCCCGCTAGCCCGGCTGCTGGGGCTGACACACAGAG 974
Db 831 GCCTCCCGAGTTTAAATTAGACCCCGCTAGCTGAGCTCTCTGGGCATCCATACCCAGAC 890
Qy 975 CCGTCTAGCCCATTTGTCAGGCGCTGTGGCAGTATGTGAAGACCAACAGGCTGCAGGACTC 1034
Db 891 TCGTCCAGTGATCATCCAAGCACTGTGCAATATATTAAAGACACATAAGCTCCAGGACCC 950
Qy 1035 CCATGACAAAGAAATACATCAATGAGGACAGTATTTCCAGCAGATTTTTTGGATGTCGCCCG 1094
Db 951 TCAGAGCGGAGTTTGTTCATCTGTGCAAGTACTCTGACAGAGATCTTTTGGTCTCAACG 1010
Qy 1095 GCTGAAGTTTTCTGAGATTTCCCGCAGCGCTCACAGCCCTGCTATTGCCCCCTGACCCCAAT 1154
Db 1011 TATGAAGTTTTCAGAGATCCCTCAGCGGCTCCATGCTCTGCTTATGCAACCAAGACCTAT 1070
Qy 1155 TGTATCAACCATGTCATCAGCGTGGACCTTTCAGACCAAGAAAGACAGCGGTGCTATGA 1214

```



```

Db      1071 CATCATTAATCATGTCATGATGTTGATACCCGAAATGATCAGAAAAAGACAGCTTTGATGA 1130
Qy      1215 CATTCAGTGAAGTGAAGAGAGCCATTAAAGGGGAGATGAGACGCTTCTCTTATCCAC 1274
Db      1131 CATGATGTTGAAGTGAAGTGAAGCCTTGAAGACCCGATGAATCTTTCTGCTGCTCCAC 1190
Qy      1275 GGCACCAACGACGAGAGATGATCTCTTGACAGTAAGATTCATGAGCAGATTGATGCAT 1334
Db      1191 TCCAGGCCAACAGAGATGTTGACTACTAGACAACAAGATTCATGAGACATGAAGAACAT 1250
Qy      1335 AAACCCGCTCAAGATTCAGAGGACTTCACTGTAAGCTTCTCCAGAGACCCCAAGGCTA 1394
Db      1251 CAACACGCTGAAGACTCAGCGGGAGTTCATGCTGAGCTTGGCAGAGACCCCTCAGGTTT 1310
Qy      1395 TGTCCAGAGCTGCTCCGCTCCAGAGCGGGAGCCCAAGCTGATGATGATGATGAGCCGG 1454
Db      1311 CATCAATGACTGCTTCACTTCACTGAGGAGGAGCCCTCAAGACAAATGATGATGATGAGG 1370
Qy      1455 CAACCCCTGAGAGAGAGCGCGGGCTGAGTTTACCAACGAGCCCTGCTCCAGAGAGCCGT 1514
Db      1371 TAACCCAGAGAGAGAGCGCGGCGGAGTGTCTAATCTCCAGCCCTGGGCTCAGAGAGCTGT 1430
Qy      1515 CAGTGGCTACTTCTTACTGCAAGATTCAGACGCGCAGGAGAGCTGAGACAGTCTGCT 1574
Db      1431 GTGCCGATCTTCTTACTCCAGGTGACGAGACGACGACAAATTAAGACCAAGCCCTGGG 1490
Qy      1575 TGTGGCGACACCTTGAAGGCCCAAAACA 1603
Db      1491 AATCCGATACATGAGGCGCTCTCCACA 1519

RESULT 7
BC009368
LOCUS      BC009368      3335 bp      mRNA      linear      PRI 06-JUN-2003
DEFINITION      Homo sapiens SWI/SNF related, matrix associated, actin dependent
                  regulator of chromatin, subfamily d, member 1, mRNA (CDNA clone
ACCESSION      MGC:15280 IMAGE:4110499), complete cds.
VERSION      BC009368
KEYWORDS      MGC.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens (human)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 3335)
Strausberg,R.L., Peirngold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D.,
Altshul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Datchenko,L., Martusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carinich,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulik,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,U., Helton,E., Kettleman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmitt,J., Myers,R.M.,
Butterfield,Y.S., Krzyzanski,M.I., Skliska,U., Smallie,D.E.,
Schneringer,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL      MEDLINE
PUBMED      22388257
12477932
TITLE      2 (bases 1 to 3335)
REFERENCE      Strausberg,R.
AUTHORS      Direct Submission
JOURNAL      Submitted (12-JUN-2001) National Institutes of Health, Mammalian
                  Gene Collection (MGC), Cancer Genomics Office, National Cancer

```

```

REMARK
COMMENT
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Latic,P., Legaapi,R.,
Maduro,Q.L., Mastello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantipop,S., Thomas,P.J., Touchman,J.W.,
Tsurgenev,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Series: IMAGE Plate: 26 Row: b Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 21261439.

FEATURES
source
1..3335
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:15280 IMAGE:4110499"
/tissue_type="muscle, rhabdomyosarcoma"
/clone_1ib="Nih MGC_17"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
1..3335
/gene="SMARCD1"
/notes="synonyms: BAF60A, CRACD1"
/db_xref="locusID:6602"
/db_xref="MIM:601735"
130..1677
/codon_start=1
/product="SMARCD1 protein"
/protein_id="AAH09368.2"
/db_xref="GI:31455235"
/db_xref="locusID:6602"
/translation="MAARAGFOSVPSGAGSAGGAGAAALPGCTPGPPVPMGPAP
GGLYRSPMGAAVYRPMGLDGSMTPGSPMGPGVGNSSVPRGLAOSGDOSRKR
PAPOIIOVOOAAVONRNHNAKKKMAADKILPQRIREVPSQAMDLAVERKLDOT
IMRKLIDIOELAKRITKOKRKLRIFISTFPKASDADGSAWSLRYEGRLEB
DASLKYDITKDKRKSFFKSLVLELDLIGPDHMLVHRHTATTOETDQVVRG
VNVAICTVLMIDYPPQFQKIDPLRLAGITQRPVILIAIMQYIKTKIDQDHER
EVIDDKYLOOIFESORMKFSEIPORLHALMPPEPIIINHVISVDPNDOKTACYDI
DVEVDITKTMNSRSLSTASQOATATIDNKTHETITINOLKQREERMLSPARPOG
PINDLQSCQCDLKTMTDVVGNPPEERRAEFTFQPMAGDAVCRFYRSKVQORQELBO
ALGIRNT"
BASE COUNT      746 a      954 c      846 g      789 t
ORIGIN
Query Match      40 1%; Score 690.6; DB 9; Length 3335;
Best Local Similarity 70.7%; Pred. No. 3,9e-121;
Matches 996; Conservative 0; Mismatches 369; Indels 44; Gaps 4;
Qy      217 CAGCCCGCCCGGATCCGCTTGAGAGCCGAGATGCCCAAGAGGGCGCCCATGAGGCCCC 276
Db      304 CCGAGACACGATGATTTGCGACGAGCCGAGATGACCTCAGGAGCCTTCCATGAGACCC 363
Qy      277 CGGGCTCCCGTACATGAGGACGCCCGCGTGCAGACCCGGGCTGGCCCCCG-GGGATG 335
Db      364 C-----CTGGCTATAGGGGGGAACCTTCAGTCCGACTGGCGCTGCCCAAGTGAAGGATGG 418

```

```
QY 336 AGCCGCGCCGCAAGCAGCAGCGCCCCCGCCGCGCAGAGCCAGAGCCAGGGCCCA 395
Db 419 ATCATGTCGCAAGAGACTGCGCCCTCAG-----CAGATCCACAGAGT 461
QY 396 GCCGAGCCACCGCCCCCGCGCAGCGCAGTGCACAGAGGAGGAGATGGCTGACAA 455
Db 462 CCAGCAGCGCGGTCCAAAATCGAAAACCAATGCAAAAGAAAAGAAAGATGGCTGACAA 521
QY 456 AATCTCTCTCAAAAGGATTCGGGAGCTGGTCCCCGAGTCCAGGCTTACATGGACCTCTT 515
Db 522 AATCTCTCTCAAAAGGATTCGGGAGCTGGTACCAGAAATCCAGGCTTATGGATCTCTT 581
QY 516 GCATTTGAGAGGAACTGGATCAAAACCATCATGCGGAAGGGGTGGACATCCAGAGGC 575
Db 582 GGCCTTTGAAAGGAACTGGACAGACTATCATGAGGAAACGGCTAGATATCCAGAGGC 641
QY 576 TCTGAAGGCCCATGAAGCAAAAGCGGAAGCTGCGACTCTATATCTCCACACTTTTAA 635
Db 642 CTTGAACAGTCCCATCAAGCAAAAACGGAAGCTGCGAATTTTCATTTCTAACACTTTCAA 701
QY 636 CCCTCGGAAGTCTGATGCTGAGGATTCGACGGCAGCATTCCTCTCGGAGCTACGGGT 695
Db 702 TCCGCTAAGTCAGATGCGGAGATGGGAAGGACGGTGGCTTCTCGGAGCTTCGGGT 761
QY 696 GGAGGGGAAGCTCTGG-----ATGATCCAGCAAAACAGAGCG 734
Db 762 AGAAGCAGGCTCTCGGAGGATTCAGCCTTGTCCAAATATGATGCCACTAAACAAAAGAG 821
QY 735 GAAGTCTCTCTTTCTTCAAGAGTTTGGTCATCGAGCTGACAAAGATCTTTATGGCCC 794
Db 822 GAAGTCTCTCTCTTTTAAAGTCTCTGGTGATGAACCTGGACAAAGACCTGTATGGGCC 881
QY 795 TGACAAACACCTCGTTGAGTGGCATCGACACCCACGACCCAGGAGACGGCTTCCA 854
Db 882 AGACAAACCATCTGGTAGAATGGCAGACACCGCCACTACCCAGGAGACCGATGGCTTCA 941
QY 855 GGTGAACCGGCTGGGGACCTGAGTGTGCGTGCACGCTGTCTCTCATCTGCTGGACTACA 914
Db 942 GGTGAACCGCGCGGAGACGTTGAATGTACGGTGTACTGCTCTACTGATGCTGGATTACA 1001
QY 915 GCCTCCACAGTTCAAACTGGATCCCGCTAGCCCGCTGCTGGGCTGCACACACAGAG 974
Db 1002 GCCTCCACAGTTTAAATTTAGACCCCGCTAGCTGCTGCTGCGCATCCATACCCAGAC 1061
QY 975 CGCTCAGCATTTGTCAGGCGCTGTGCAGTATGTGAAGACCAACAGGCTGCAGGACTC 1034
Db 1062 TCGTCCAGTGATCATCCAGCACTGTGCAATATATTAAAGACATAAGCTCCAGGACC 1121
QY 1035 CCATGACAAGGAATACATCAATGGGGAACAAGTATTTCCAGCAGATTTTGTGATTTGCCCG 1094
Db 1122 TCACGAGCGGAGTTTGTTCATCTGTGACAAGTACCTGCAGCAGATCTTTTGAGTCTCAACG 1181
QY 1095 GCTGAAGTTTCTGAGATTTCCAGCGCTCACAGCCTCTGCTATTGCCCCCTGACCCAT 1154
Db 1182 TATGAAGTTTTCAGAGATCCCTCAGCGCTCCATGCTTGTCTTATGCCCCAGAACCTAT 1241
QY 1155 TGTCAATCAACATGTCATCAGCTGACCCCTTCAGACCAAGAGAGACAGCGTGTATGA 1214
Db 1242 CATCATTAATCATGTTCATCAGTTGTGACCCGAATGATCAGAAAAGACAGCTTGTATGA 1301
QY 1215 CATTGACGTGGAGTGGAGGACCAATTAAGGGGCGAGATGAGCAGCTTCTCTCTATCCAC 1274
Db 1302 CATTGATGTGAAGTGGATGACACCTTGAAGACCCAGATGAATCTTTCTCTGCTGCTCAC 1361
QY 1275 GGCAACACAGAGGAGATCAGTCTCTCGACAGTGAAGTATCATGAGACGATTTGATCCAT 1334
Db 1362 TGCCAGCAACAGGAGATTTGCTACTCTAGACAAACAAAGATCCATGAGACAAATAGAAACCAT 1421
QY 1335 AAACAGGCTCAAGATCCAGAGGAGCTTCATGCTAAGCTTCTCCAGACACCCCAAGGGCTA 1394
Db 1422 CAACAGCTGAAGACTCAGCGGAGTTTCATGCTGAGCTTTGCCAGAGACCTTCAGGGTTT 1481
QY 1395 TGTCCAAGACCTGTCTCCGCTCCAGAGCCGGACCTCAAGGTGATGACAGATGTAGCCGG 1454
```

```
Db 1482 CATCAATGACTGGCTTCAGTCCCAGTCCAGGACCTCAAGACAATGACTGATGTGGTGG 1541
QY 1455 CAACCCCTGAAGAGAGCGCGGCTGAGTTCTACACAGCCCTGGTCCCAAGGAGCGCT 1514
Db 1542 TAACCCAGAGAGGAGCGCGAGCTGAGTTCTACTTCAGCCCTGGGCTCAGGAGGCTGT 1601
QY 1515 CAGTCGTACTTCTACTGCAAGATCCAGACGCGCAGCAGGAGCTGGAGCAGTCCGTGGT 1574
Db 1602 GTGCCGATACTTCTACTCCAAAGTGCAGCAGACGACAAGAAATTAGAGCAAGCCCTGGG 1661
QY 1575 TGTGCGCAACACCTAGGAGCCCAAAACA 1603
Db 1662 AATCCGAATACATAGGCGCTCTCCCAACA 1690

RESULT 8
AX305314
LOCUS AX305314 3103 bp DNA linear PAT 11-DEC-2001
DEFINITION Sequence 65 from Patent WO0188188.
ACCESSION AX305314
VERSION AX305314.1 GI:17644884
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Ishikawa,K., Asai,S., Takahashi,Y., Nagata,T. and Ishii,Y.
TITLE Method for examining ischemic conditions
JOURNAL Patent: WO 0188188-A 65 22-NOV-2001;
School Juridical Person Nihon University (JP)
FEATURES
Location/Qualifiers
1..3103
source
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
BASE COUNT 657 a 925 c 817 g 704 t
ORIGIN

Query Match 39.7%; Score 685.2; DB 6; Length 3103;
Best Local Similarity 71.1%; Pred. No. 4.1e-120;
Matches 997; Conservative 0; Mismatches 358; Indels 47; Gaps 5;

QY 217 CAGCGCCCGCGGATGCGGTCTTGAGCGCGGATGCCACAGGCGCGCCCATGGCCCCC 276
Db 166 CCGAGACCGAGTATGCTGCCAGGTAGCCGATGACACTCAGGACCTTCATGGGACT 225
QY 277 CGGCTCCCCGTACATGGGACGCCCGCGTGCACCCCGCTGGCCCCCGC-GGGATGG 335
Db 226 C-----CTGGTATGGGGGAAACCTTCAGTCCGACCTGTGCTGGCCAGTCAGGGATGG 280
QY 336 AGCCCGCCCGCAGGAGCAGCGCCCCCGCCGCGCAGCGCAGCCAGCAGACGAGGCGCA 395
Db 281 ACCAGTCCCGCAAGAGACCTGCACCTC-----AACAGATCCACAGGT 323
QY 396 GCCGAGCGCCACCCCGCGCGAGCGCGAGTGCAGAGGAGGAGAGATGGCTGACAA 455
Db 324 CCAGCAGCGCGGTCCAAAATCGAAATCAATGCAAGAAAAGAAAGATGGCTGACAA 383
QY 456 AATCCTCCTCAAAAGGATTCGGGAGCTGGTCCCCGAGTCCCGAGCTTACATGGACCTCTT 515
Db 384 AATCCTACCTCAAAAGGATTCGGGAACTGGTCCCAAGAAATCAAGGCTACATGGATCTCT 443
QY 516 GGCAATTTGAGAGGAACTGGATCAACCATCATCGGAAGGGGTGGACATCCAGGAGGC 575
Db 444 GGCTTTTGAAGGAAACTGGACCACTATTATGAGGAAGCGGTAGATATCCAGGAGGC 503
QY 576 TCTCAAGAGGCCCATGAAGCAAAAGCGGAAGCTCGGACTCTATATCTCCAACTTTTAA 635
Db 504 CTTGAACGCTCCCATCAAGCAAAAACCGGAAGCTCGGAATTTTCATTTCTAACACGTTCAA 563
QY 636 CCCTGCGAAGTCTGATGCTGAGGATTCGACGGCAGCAATTCCTCTCTGGGAGCTACGGGT 695
```



```
Qy 456 AATCTCTCCCTCAAAGGATTCGGAGCTGGTCCCGAGTCCAGGCTTACATGGACCTTT 515
    |||||
Db 384 AATCTCTACTCAAGGATTCGGAACTGGTCCAGAAATCAAGGCTTACATGGATCTCT 443
    |||||
Qy 516 GGCATTTGAGAGGAACTGGATCAAAACATCATGCGGAAGGGGTGGACATCCAGGAGGC 575
    |||||
Db 444 GGCATTTGAAAGGAACTGGACAGACTATTATGAGGAAGGGCTAGATATCCAGGAGGC 503
    |||||
Qy 576 TCTGAGAGGCCCATGAAGCAAAAGCGGAAGCTCGACTCTATATCTCAACACATTTTAA 635
    |||||
Db 504 CTTGAAACGTCCTCAAGCAAAACGGAAGCTCGGAATTTTCATTTCTAAACGCTTCAA 563
    |||||
Qy 636 CCTCGGAAGTCTGATGCTGAGGATTCGAGCGGACGATGCTCTCTGGGAGCTACGGGT 695
    |||||
Db 564 TCCGCTTAAGTCGAGCGGAGGATGGGGAAGGACGGTGGCTTCTCTGGGAGCTCCGGT 623
    |||||
Qy 696 GGAGGGGAAGCTCTGG-----ATGATCCAGCAACAAAGAGCG 734
    |||||
Db 624 AGAAGGCGGCTCCTGGAGGACCGCGCTTGTCAAAATATGACGCCACCAAGCAAAAGAG 683
    |||||
Qy 735 GAAATCTCTCTCTCTCAAGAGTTTGGTTCATCGAGCTGGACAAAGATCTTTATGGCCC 794
    |||||
Db 684 AAGATCTCTCTCTCTTTTAAAGTCTCTGGTGAATCGAATGGACAAAGACCTCTATGGCCC 743
    |||||
Qy 795 TGACAACACCTCGTTGAGTGGCATCGGACACCCAGCAGGAGCGGACGGCTTCCA 854
    |||||
Db 744 AGACAAACCATGTGTAGATGGCAGACACCGCCACTACCCAGGAGCGGATGGCTTCCA 803
    |||||
Qy 855 GGTGAAACGGCTGGGGACCTGAGTGTGCGTGCACGCTGTCTCTCATGCTGGACTACA 914
    |||||
Db 804 GGTGAAGCGGCCAGGAGATGTGAATGTACGGTGTACTGTCTCTGTGATGCTGGACTACA 863
    |||||
Qy 915 GCCTCCCAAGTCAAACTGGATCCCGCTAGCCCGCTGCTGGGGCTGCACACACAG 974
    |||||
Db 864 GCGCCCGGAGTTAAATTAGACCTCGCTCGCTCTTGGGCACTCCATACCCAGAC 923
    |||||
Qy 975 CGCTCAGCATTTGTCAGGCGCTGTGGCAGTGTGTAAGACCAACAGCGCTGCAGACTC 1034
    |||||
Db 924 AGCTCAGTGATCATCAAGCACTGTGGCAGTATATTAAACACACAGCTCCAGGACCC 983
    |||||
Qy 1035 CCAAGCAAGGAATACATCAATGGGGAACAAGTATTTCCAGCAGATTTTGTGTCCTCCG 1094
    |||||
Db 984 TCACGAGCGAGATTTGTCTCTGTGACAAGTACCTCCAGCAGATCTTTGAATCTCAGCG 1043
    |||||
Qy 1095 GCTGAAGTTTCTGAGATTCGCCAGCGCTCAACGCGCTGTATGCGCCCTGACCCCAAT 1154
    |||||
Db 1044 GATGAAGTTCTCAGAGATCCCTCAGCGCTCCACGCTTGTCTTATGCCACAGGCCCCAT 1103
    |||||
Qy 1155 TGTCTATCAACCATGTCTCAGCGTGGACCTTTCAGACCAAGAAAGACAGCGGTGCTATGA 1214
    |||||
Db 1104 CATCATCAATCATGTCTCAGTGTGGACCCCAATATGACCAAGAAAGACCGGTGCTATGA 1163
    |||||
Qy 1215 CATTGACGTGGAGTGGAGGACCAATTAAGGGGAGATGAGCAGCTCTCTCTTATCCAC 1274
    |||||
Db 1164 CATTGACGTGGAGTGGATGACACTCTGAAGACCCAGATGAATCTTCTCTGTTGTCAC 1223
    |||||
Qy 1275 GGCACACAGCAGGAGATCAGTCTCTGACAGTGAAGTCCATGAGACGATGAGTTCAT 1334
    |||||
Db 1224 TGCCAGCCAGGAGATCGCCACTCTAGACCAAGAGATCCATGAGACGATAGAGACCAT 1283
    |||||
Qy 1335 AAACAGCTCAAGATTCAGAGGAGCTTCATGCTAAGCTTCTCCAGACACCCCAAGGCTA 1394
    |||||
Db 1284 CAACAGCTGAAGACCCAGCAGAGATTTCATGTTGAGCTTGCCC---GAGCCCTCAGGTTT 1340
    |||||
Qy 1395 TGTCAAGACCTGTCTCCGCTCCAGAGCGGACCTCAAGGTGATGACAGATGTAGCCGG 1454
    |||||
Db 1341 CATCAATGATTGGCTTTCAGTCCCGCAGTGGGCGCTCAAGACGATGACTGTTGTTGGG 1400
    |||||
Qy 1455 CAACCTGAAGGAGCGCGGCTGAGTTCTTACACACGCGCTGCTCCAGGAGCGCT 1514
    |||||
Db 1401 TAACCCGGAAGAGGAGCGCTGCTGCTGAGTTCTTACTTCAGGCGCTGGCTCAGAGGCTGT 1460
    |||||
```

```
Qy 1515 CAGTCGTACTTCTACTGCAAGATCCAGCAGCGCAGGAGAGCTGGAGCAGTGCCTGGT 1574
    |||||
Db 1461 GTGCCGATACTTCTACTCCAAAGGTGCAGCAGGCGCGCAAGAGTTAGAGCAAGCCTGGG 1520
    |||||
Qy 1575 TGTGCGCAACACCTAGGAGCCC 1596
    |||||
Db 1521 AATCCGAACACATAGGCGCTC 1542
    |||||

RESULT 10
BC049347
LOCUS BC049347 2747 bp mRNA linear VRT 31-MAR-2003
DEFINITION Danio rerio, clone MGC:55436 IMAGE:2639578, mRNA, complete cds.
ACCESSION BC049347
VERSION BC049347.1 GI:29387083
KEYWORDS MGC.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 2747)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: gcapsb-remail.nih.gov
Tissue Procurement: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
cDNA Library Preparation: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 101 Row: 1 Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein.

FEATURES             Location/Qualifiers
     1..2747
         /organism="Danio rerio"
         /mol_type="mRNA"
         /strain="AB"
         /db_xref="taxon:7955"
         /clone="MGC:55436 IMAGE:2639578"
         /tissue_type="Whole body, adult, (one male and one female,
         including unfertilized eggs)"
         /clone_lib="Sugano Kawakami zebrafish DR"
         /lab_host="DH10B"
         /note="Vector: pME18S-FL3"
     28..1560
         /codon_start=1
         /product="Unknown (protein for MGC:55436)"
         /protein_id="AAH49347.1"
         /db_xref="GI:29387084"
         /translation="MAARSGFQPSPTGGGAPMGPPVAGAPMGPGTSGRMGPG
         GPQNHMTSRMPGYPFRGMPMPADRMTPQGGSPGPGYCASVPSPGMPVMPDSRKP
         PNOIQVQQNRNOHAKKWKADILPQRIRELVEPQAYMDLLAFERKLDQTNMRK
         LDIOEALRKIFQKRLRIFISNTFNPAKPADEGTVASWELRVEGLDITAVSK
         YEATQKRKSFSSFKSLVIDLQYGDNLHVEVHTATQETDGFQVKGPDGVGR
         CTVLLMDYDPOPKVLDLRLALLHTQRPVLIQALWQVVKLQDPHEREFINC
         DKYLQIQFEQTKMFSIPORLHALLMPPBPIIINHVSVDNDQKTKTCDVIDEVD
         DTLKQNSFLLSTASQOETAGLNDKTHETIETINQLKTQREFMLSPFARDPQGFNDW
         LQSCDRCLTKMTDVGVPBERRAEFYIQPWAQEAFCVRYFYSKVQQRQEQLEALGIR
         NT"
```



Matches 996; Conservative 0; Mismatches 357; Indels 49; Gaps 7;

QY 217 CAGGCCCGGGATGCGCTCTGGAGCCGCGATGCCACACAGGGGCGCCATGGGCCCCC 276  
Db 166 CCGAGACAGATGCTGCCAGGTAGCGAATGACACCTCAGGACCTTCCATGGACCT 225  
QY 277 CGGGCTCCCGTACATGGGACGCCCGCGTGCAGCCCGGCTGGCCCCCGC-GGGATGG 335  
Db 226 C-----CTGGCTATGGGGGAACCTTCAGTCCGACCTGGTCTGGCCCACTCAGGGATG 280  
QY 336 AGCCCGCCGCAAGGACAGGCCCGCCCGCGCAGACCCAGGCACAGAGCCAGGGCCA 395  
Db 281 ACCAGTCCCGCAAGAGACCTGCACCTC-----AACAGATCCAGCAGGT 323  
QY 396 CGCGAGCCCAACCGCCCGCGGAGCGCGAGTCCCAAGAGGAGAGATGGCTGACAA 455  
Db 324 CCAGCAGCAGGGGTCCAAATCGAATCAATGCAAGAAAGAAAGATGGCTGACAA 383  
QY 456 AATCCTCCTCAAGGATTCGGGAGCTGGTCCCGAGTCCAGGCTTACATGGACCTTT 515  
Db 384 AATCCTACCTCAAGGATTCGGGAATCGTCCAGAATCACAGGCTACATGGATCTCT 443  
QY 516 GGCATTGAGAGGAACCTGGATCAAAACCATCATGCGGAAGGGGTGGACATCCAGGAGC 575  
Db 444 GGTCTTTGAAAGGAACCTGGACCACTATTATGAGGAAGCGGTAGATATCCAGGAGGC 503  
QY 576 TCTGAAGAGGCCATGAAGCAAAAGCGAAGCTGCGACTCTATATCTCCAAACACTTTAA 635  
Db 504 CTTGAAACGTCCCATCAAGCAAAACGGAAGCTGCGAATTTTCATTTCTAACAGTTCAA 563  
QY 636 CCCTCGAAGTCTGATGCTGAGGATTCGCGAGGAGCATTCCTCTCGGAGCTACGGGT 695  
Db 564 TCCGGCTAAGTCGAGCGCGGAGGATGGGAGGAGCGGTGGCTTCTCGGAGCTCCGGGT 623  
QY 696 GAGGGGAAGCTCTGG-----ATGATCCCAACCAAGAGCG 734  
Db 624 AGAAGCGCGCTCTCGGAGGAGCGCGCTTGTCCAATATAGCGCCACCAAGCAAAAG 683  
QY 735 GAAGTCTCTCTTTCTTCAAGAGTTTGGTTCATCGAGCTGGCAAGATCTTTATGGCC 794  
Db 684 AAGTTCTCTCTTTTAAAGTCTTGGTGTATGGAATCGAATCGAACAAGACCTCTATGGCC 743  
QY 795 TGACAACCACTCTGTAGTGGCATCGGACACCCACGACCCAGGAGACGGCTTCCA 854  
Db 744 AGACAACCATCTGTGTAGATGSCACAGACCGCCACTACCCAGGAGACCGATGGCTCCA 803  
QY 855 GGTGAACCGCTGGGGACCTGATGTGCGCTGCAGCTGTCTCTATCTGCTGGACTACA 914  
Db 804 GGTGAAGCGCGCAGGAGATGTGAATGTACGGTGTACTGTCTGTCTGATGCTGGACTACA 863  
QY 915 GCCTCCCGAGTTCAAACTGGATCCCGCTAGCCCGCTGTCTGGGGCTGCACACACAGAG 974  
Db 864 GCCCGCCAGTTTAAATTAGACCTTCGCTGCTGCTGGCTCTTGGGCATCCATACCCAGAC 923  
QY 975 CCGCTCAGCCATTTGTCAGGCCCTGTGGCAGTATGTGAAGCAACACAGGCTGCAGGACTC 1034  
Db 924 ACGTCCAGTGATCATCCAGCACTGTGCGAGTATATTAACACACAAAGCTCCAGGACCC 983  
QY 1035 CCATGACAAGGATACATCAATGGGGACAAGTATTTCCAGCAGATTTTGTATGTCCCG 1094  
Db 984 TCACGAGCGAGAGTTTGTCTGTGTGACAAAGTACCTCCAGCAGATCTTTGAATCTCAGCG 1043  
QY 1095 GCTGAAGTTTCTGAGATTCCCGACGCTCACACGCTGTATTTGCCCCCTGACCCAAT 1154  
Db 1044 GATGAAGTTCTCAGAGATCCCTCAGCGGTCCACGCTTGTATGTCACACAGGCCAT 1103  
QY 1155 TGTCTAACCATGTCTCAGCGTGGACCTTTCAGACCAGAAAGACAGCGGTGCTATGA 1214  
Db 1104 CATCATCAATCATGTCTCAGTGTGGACCCCAATGACCAGAAAAAGACCGCGTGTATGA 1163  
QY 1215 CATTGACGTGGAGTGGAGGCCATTAAAGGGGAGATGAGCAGCTTCTCTTATCCAC 1274  
Db 1164 CATTGACGTGGAGTGGATGACACTCTGAAGACCCAGATGAACCTCTTTCTGTTGTCAC 1223

QY 1275 GGCAACACAGCAGGAGATCAGTCTCTGGACAGTAAGATCCATGAGACGATTGAGTCCAT 1334  
Db 1224 TGCCAGCAGCAGGAGATCGCACTCTAGACAACAAGATCCATGAGAGATAGAGACCAT 1283  
QY 1335 AAACACAGCTCAAGATCCAGAGGACCTTCATGCTTAAGCTTCTCCAGAGACCCCAAGGCTA 1394  
Db 1284 CAACCACTGAAGACCCAGCAGAGATTCATGTTGAGCTTGCCC---GAGCCCTCAGGTTT 1340  
QY 1395 TGTCAAGACCTGCTCGCTCCAGAGCGGAGCTCAAGGTGATGACAGATGTAGCCGG 1454  
Db 1341 CATCAATGATTGGCTTCAGTCCAGTCAGGG---CTCAAGACGATGACTGATGTGGTGG 1399  
QY 1455 CAACCTCTGAAGAGAGCGCCGGCTGAGTTCTACACACAGCCCTGTCACAGGAGCGCT 1514  
Db 1400 TAA---CCGGAAGAGAGCGCTGCTGCTGAGTTCTACTCCAGCCCTGGCTCAGGAGGCTGT 1458  
QY 1515 CAGTCGCTACTTCTACTGCAAGATCCAGCAGCAGCAGGAGCTGGAGCAGTCTGCTGGT 1574  
Db 1459 GTGCCGATACTTCTACTCCAAGGTGCAGCAGAGCGCGCAAGAGTTAGAGCAAGCCCTGGG 1518  
QY 1575 TGTGCGCAACACCTAGGAGCCC 1596  
Db 1519 AATCCGAACACATAGGCGCTC 1540

RESULT 12  
AB003504  
LOCUS 2297 bp mRNA linear ROD 05-FEB-1999  
DEFINITION Rattus norvegicus mRNA for BAF60b, partial cds.  
ACCESSION AB003504  
VERSION AB003504.1 GI:2723483  
KEYWORDS BAF60b.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (sites)  
AUTHORS Nomoto, K., Nakazato, S., Kazahari, K. and Ono, M.  
TITLE Gene structure of rat BAF60b, a component of mammalian SW1/SNF complexes, and its physical linkage to the growth hormone gene and transcription factor SUG/proteasome p45 gene  
JOURNAL Gene 202 (1-2), 157-165 (1997)  
MEDLINE 98087432  
PUBMED 9427560  
REFERENCE 2 (bases 1 to 2297)  
AUTHORS Ono, M.  
TITLE Direct Submission  
JOURNAL Submitted (02-MAY-1997) Masao Ono, Rikkyo University, College of Science, Department of Chemistry, Life Science Course; Nishi-Ikebukuro, 3-34-1, Toshima-ku, Tokyo 171-8501, Japan (E-mail: monoo@rikkyo.ac.jp, Tel:+81-3-3985-2387, Fax:+81-3-5992-3434)  
COMMENT Rat BAF60b is closely related to human BAF60b (Genes & Dev. 10, 2117-2130, 1996).  
FEATURES Location/Qualifiers  
source 1. 2297  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Wistar"  
/db\_xref="taxon:10116"  
/clone="HL2"  
/sex="male"  
/tissue\_type="liver"  
c1..1411  
/codon\_start=2  
/product="BAF60b"  
/protein\_id="BAA24105.1"  
/db\_xref="GI:2723484"  
/translation="GPAAPAAQYQPCGSRMPMAGLVQVPPAGSPFGTAAPLRPG MPPTMDPPFRKLLVLPQAPQPMQPKRKLKRVADKVLPORIELVPEQAYMDLLA FERKLDQTIARKRMEIQEATKKPLTKRKLRIYISNTFSPSKADGDSGTAGTGGTT

polyA\_signal  
BASE COUNT 522 a 624 c 623 g 528 t  
ORIGIN  
AADKVASWELRVGKLLDDPSKOKRFFSFKSLVLELDKELVGPNNHLEWHMPTT  
OETDPOVRKPDGLNVKCTLLMLDHPOLYKLDPLRLHGTOTRAIMQALPLY  
IKNLOADHREERYINONRYFOIIFSCGRLREIIPMKJALQHPDPIVINHVLSD  
PNDOKTACTDIDIVEDDPLKOMSFLASTTNOQISALDKIHETIISINOKTOR  
DFMLSTFEPQDITQEMLSORRDLKIIDVIGNPEERRAAFYHOPMAQEAVERHIF  
AVQQRORLEOLYGLRLT"

Query Match 38.0%; Score 655.4; DB 10; Length 2297;  
Best Local Similarity 71.8%; Pred. No. 1.9e-114;  
Matches 888; Conservative 0; Mismatches 316; Indels 33; Gaps 1;

394 CAGCCGAGACCCCGCCCGCCGAGCCGAGTCCCAAGAGGAGAGATGGCTGAC 453  
182 CAGGCCAGACCCCGCGATGCTCCGCGCCGAGGGTTAAAGAGGAGAGATGGAGAT 241  
454 AAAATCTCTCTCAAAAGATCGGAGAGTGTCCCGAGTCCGAGCTTACATGACATC 513  
242 AAGGTTCTACTCAGGAATCCGGAGCTTGTCCAGAGTCTCAGCATACATGATCTT 301  
514 TTGGCATTTGAGAGAACTGATCAACCATCATGCGAGAGGGGTGAGCATCCAGAG 573  
302 TTAGCTTTGAGAGGAAAGCTGAGCAGACCATCGCTCGCAAGGGATGAGATCCAAAG 361  
574 GCTCTGAAAGAGCCCATGAAAGCAAAAGCGGAGCTCGACTCTATATCTCCAACTTTT 633  
362 GCATCTMAAGAGCTCTGAGCGAAAGCAAACTTCGATCTATTTTCCATATCATTC 421  
634 AACCCGCGAAGCTGATGCTGAGAT----- 661  
422 AGTCCAGAGAGCAGATGTGTATTAATTCAAGACTCAGAGAGCCCTGGGGAGACCCA 481  
662 -CCGAGCGAGCATTCCTCTCTGAGAGCTACGGGTGAGGGAGAGCTCTGTGATGATCC 720  
482 GAGAGCAGACAGAGGTGCTCTCTGAGAGCTTCGAGTAAAGGAAACTTCTGATGATCT 541  
721 AGCAACAGAGAGGAGATTTCTTTTCTTAAAGATTGGTCAATCGAGCTGAGCAAA 780  
542 AGCAACAGAGAGGAGATTTCTCTTCTTAAAGCTTGTATTCAGATTGAGCAAG 601  
781 GATCTTTATGAGCCCTGACCAACCTCGTTGAGTGCATCGACACCCAGACCCAGAG 840  
602 GAGCTGTATGAGCTTACCAACCTCTGAGAGTGCATCGAGTCCACCATCTGAGAA 661  
841 AGCGAGCGCTTCCAGGTGAAACGGCTGGGAGCTGAGTGGCGCTGACGCTGCTCTC 900  
662 ACAGATGGCTTTCAGGTGAAACGGCAGAGATCTCAATGTCAAGTGCACCTCTGCTC 721  
901 ATGCTGAGTACCAAGCTTCCCAAGTTCAAACTGGATCCCGCTAGCCCGGCTGCGG 960  
722 ATGCTGAGTACCAAGCTTCCCAAGTTCAAACTGGATCCCGCTAGCCCGGCTGCGG 781  
961 CTGCAACAGAGAGCGGCTCAGCATATTCAGAGCCCTGAGAGATGTGAAGACCAAC 1020  
782 GTGCAACAGAGAGCGGCTCAGCATATTCAGAGCCCTGAGAGATGTGAAGACCAAC 841  
1021 AGGCTCAGAGAGCTCCATGCAAGAAATCATGAGGAGGAGCAAGATTTCCAGCAT 1080  
842 CAGCTGCAAGATGGAGATGAGCGGAGTACATCACTGCAACCGCATCTTCCCGCAT 901  
1081 TTTGATGTGTCCTCGGCTGAAAGTTTGTAGATTTCCCAAGGCTCAAGCGCTGTAATG 1140  
902 TTCAGTGTGCGCGCTCGTCTCTCCAGATCCCATGAAGCTGCGTGTGCGAG 961  
1141 CCCCCGAGACCAATTTCAATCAACATGATCAGAGTGAAGCCCTTCAACAGAGAGAG 1200  
962 CATTCAGAGACCAATTTTATTAACATGATTAAGTGAAGCCCTTATATAGCAAGAGAG 1021  
1201 ACAGCGTGTATGATGATGAGTGAAGTGAAGAGCCATTAAGGGGAGATGAGCAGC 1260  
1022 ACAGCGTGTATGATGATGAGTGAAGTGAAGAGCCCTTGAAGGCCAGAGATGATTAAC 1081

1261 TTCTCTCTATTCACGCGCCCAACAGAGAGATCAAGTCTCTGAGCAGTAAATCCATGAG 1320  
1082 TTCTCTGAGCTTACCAACCAACAGAGAGATTTGCTCTTGTATGTCAAGATCCATGAG 1141  
1321 ACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380  
1142 ACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1201  
1381 GACCCCAAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440  
1202 GAGCCCAAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1261  
1441 ACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500  
1262 ACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1321  
1501 TCCCAAGAGAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560  
1322 GCTCAG 1381  
1561 GACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1597  
1382 GACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1418

RESULT 13  
BC005732  
LOCUS  
DEFINITION  
Mus musculus SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 2, mRNA (cDNA clone MGC:6790 IMAGE:2647494), complete cds.  
BC005732  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 2379)  
REFERENCE  
AUTHORS  
Klausner, R.D., Collins, F.S., Wagner, C.M., Schaefer, C.F., Bhat, N.K., Altschul, S.F., Zeeberg, B., Buerow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Scheetz, T.E., Brownstein, M.J., Uedini, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullen, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shewchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Krzywicki, M.I., Skalska, U., Smal, D.E., Scherch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16895-16903 (2002)  
22388257  
MEDLINE  
PUBMED  
12477932  
2 (bases 1 to 2379)  
REFERENCE  
AUTHORS  
Strausberg, R.  
TITLE  
JOURNAL  
Submitted (02-Apr-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK  
COMMENT  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)



Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amg@bcm.tmc.edu](mailto:amg@bcm.tmc.edu)  
Gunsaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounsgood, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.

Clone distribution: MEC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 5 Row: c Column: 19  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13994210.  
 Location/Qualifiers  
 1..2379  
 source

```

/organism="Mus musculus"
/mol_type="mrna"
/strain="FVB/N"
/db_xref="taxon:10090"
/cid="MGC:26790 IMAGE:267494"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."

```

```
/CONE_IDB="NCI_CGAP_Mat11"
/lab_host="DH10E"
```

```

gene
  /note= vector: pCMV-SF0X19
  1. .2379

```

```
/gene= SmarCa2
/note="synonym: Baf60b"
```

```

/db_xref="MGI:1933621"

```

```
/codon_start=1
/product="SWT/SNP related
matrix associated actin
```

RESULT 14

HSU66618  
LOCUS

### DEFINITION





DEFINITION	Xenopus laevis, similar to SWI/SNF related, clone IMAGE:5542448, mRNA, partial cds.
ACCESSION	BC045009
KEYWORDS	BC045009.1 GI:28278759
SOURCE	Xenopus laevis (African clawed frog)
ORGANISM	Xenopus laevis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus.
REFERENCE	1 (bases 1 to 2557)
AUTHORS	Klein, S. and Strausberg, R.
TITLE	Direct Submission
JOURNAL	Submitted (23-JAN-2003) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA
REMARK	NIH-MGC Project
COMMENT	Contact: XGC help desk Email: gcapbs-remail.nih.gov Tissue Procurement: Dr. Igor Dawid cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@cgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lilias Prabhu, Parvaneh Saedi, Jacqueline Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott, Michael Thorne, Mirandaa Tsar, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAC Plate: 96 Row: C Column: 22 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein. Location/Qualifiers 1..2557

[illegible]

Db	1310	GGTAATCTGAAGAGACGTCGAGCTGATTTTACTTCCAAACATGGGCTCATGAGGCT	1369
Qy	1513	GTCAGTCGCTACTTCTACTGCAAGATCCAGACGCGCAGGAGAGCTGGAGCAGTCGCTG	1572
Db	1370	GTAATGCCGATATTTCTACTCCAAAGTGCAGCAAAAGACGTCAAGAGCTGGAACAGGCTCTC	1429
Qy	1573	GTTGTGCGCAACACCTA	1589
Db	1430	GGGATCCGAAATACATA	1446

Search completed: October 11, 2003, 10:35:12  
Job time : 4260 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 08:22:11 ; Search time 360 Seconds  
(without alignments)  
12927.309 Million cell updates/sec

Title: U66619  
Perfect score: 1724  
Sequence: 1 GAATTCGCGCGAGCGGCC.....GTCATTCATCTCGAATTC 1724

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_19jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1331	77.2	1452	25	AAD53119 Human BRG1-associ
2	1322.6	76.7	3089	22	AAH98221 Human EST-derived
3	1322.6	76.7	3272	22	AAH98319 Human EST-derived
4	963	57.0	1970	24	ABG54477 Human ovarian anti
5	821.6	47.7	6809	25	ABZ35936 Human secretory po
6	685.2	39.7	3103	24	AB199242 Mouse ischaemic co
7	642	37.2	2511	24	AB573872 Human CDNA encodin
8	572.6	33.2	2010	23	ABL02583 Drosophila melanog

9	511.4	29.7	2581	25	ABX34564 Human mdtl cDNA SE
10	504.2	29.2	4138	23	ABL02582 Drosophila melanog
11	466.2	27.0	755	25	AAD53120 Human BRG1-associ
12	410.4	23.8	450	24	ABT07117 Human ovarian can
13	410.4	23.8	450	25	ABX72995 Human ovarian carc
14	206.2	12.0	285	25	ABX47225 Bovine EST associa
15	205	11.9	452	21	AAZ80712 Human colon cancer
16	179.2	10.4	690	21	AAZ80561 Human colon cancer
17	170.4	9.9	8788	22	AA135930 Human muscleoskele
18	170.4	9.9	8788	22	AA162919 Human genomic DNA
19	170.4	9.9	8788	25	ABX58918 cDNA encoding nove
20	167.2	9.7	187	16	AAT25158 Human gene signalu
21	137.8	8.0	310	22	AAF79989 Nucleotide sequenc
22	132	7.7	520	22	ABA22582 Probe #8048 for ge
23	132	7.7	520	22	AAK10561 Human brain expres
24	132	7.7	520	22	AA117316 Probe #7249 for ge
25	132	7.7	520	22	AA142206 Probe #10892 used
26	132	7.7	520	24	ABSI0458 Human genome-deriv
27	129.2	7.5	219	22	ABA39473 Human brain expres
28	129.2	7.5	219	22	AAK23257 Probe #16469 for g
29	129.2	7.5	219	22	AA126536 Probe #23976 used
30	129.2	7.5	219	22	AA155290 Human genome-deriv
31	129.2	7.5	219	24	ABX22909 BAC containing rep
32	110	6.4	95223	21	AA122282 Drosophila melanog
33	95.6	5.5	3824	23	ABL03782 Human nervous syst
34	91.6	5.3	15500	22	ABAI5840 Human cardiovascular
35	91.6	5.3	15500	22	AA316609 Secreted protein g
36	91.6	5.3	15500	25	ABZ73784 Human secreted pro
37	91.6	5.3	15500	25	ABT16889 Human breast and o
38	91.6	5.3	15500	25	ABZ67378 Human ovarian can
39	84.6	4.9	2017	21	AAF21750 Arabidopsis thalia
40	67.8	3.9	215	24	ABL65508 Human adenosine A1
41	66.2	3.8	740	24	ABO5483 Human spliced tran
42	64	3.7	11955	20	ABN33431 Human adenosine A1
43	60	3.5	60	24	AAK53491 Breast cancer mark
44	59.4	3.4	11955	20	
45	57.2	3.3	335	25	ABT21717

## ALIGNMENTS

## RESULT 1

AAD53119 standard; DNA; 1452 BP.

AC AAD53119;

DT 28-MAY-2003 (first entry)

DE Human BRG1-associated factor (BAF) 60c2 DNA.

XX Human; nuclear receptor cofactor; proliferator-activated receptor; RXR;

XX PPAR-gamma; retinoid X receptor; cell proliferative disorder; infectious;

XX metabolic disorder; cardiovascular disorder; inflammatory disease; acne;

XX neurodegenerative disease; Parkinson's disease; psoriasis; cancer; HIV;

XX renal disease; atherosclerosis; Alzheimer's disease; diabetes; vaccine;

XX osteoporosis; human immunodeficiency virus; BRG1-associated factor; BAF;

XX gene therapy; gene; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..1452 /tag= a

FT /product= "BAF60c2 protein"

PN MO200294877-A2.

PD 28-NOV-2002.

PF 23-MAY-2002; 2002WO-1B02939.

PR 23-MAY-2001; 2001US-292526P.  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX Auwerx J, Gelman L, Debril M, Rocchi S, Picard F;  
XX WPI; 2003-120789/11.  
DR P-PSDB; AAE34718.  
XX New human PPAR or RXR cofactor polypeptide, useful for identifying  
PT modulators of nuclear receptor activity, and for treating or preventing  
PT pathologic conditions associated with cell types that express PPAR  
PT receptors, e.g. cancer  
XX  
PS Example 1; Column 55; 31pp; English.  
XX  
CC The invention relates to human nuclear receptor cofactor polypeptides.  
CC The invention particularly relates to proliferator-activated receptor  
CC (PPAR)-gamma and retinoid X receptor (RXR) cofactor polypeptides.  
CC Polypeptides of the invention are useful for screening compounds that  
CC modulate the interaction of the nuclear receptor with the nuclear  
CC receptor cofactor. Modulators of the invention are useful in preparing  
CC a pharmaceutical composition for treating and/or preventing diseases  
CC or pathologic conditions associated with cell types that express PPAR  
CC receptors. The pathologic conditions treated include metabolic or cell  
CC proliferative disorders such as diabetes, cardiovascular diseases  
CC (e.g. atherosclerosis), renal diseases, neurodegenerative diseases  
CC (e.g. Parkinson's disease, Alzheimer's disease), inflammatory diseases  
CC (e.g. psoriasis, acne), wounds, osteoporosis, infections (e.g. HIV)  
CC or cancer. The invention is useful in gene therapy and as vaccines.  
CC The present sequence is human BRG1-associated factor (BAF) 60c2 DNA.  
CC This sequence is used in the exemplification of the invention.  
XX  
SQ Sequence 1452 BP; 330 A; 447 C; 424 G; 251 T; 0 other;  
  
Query Match 77.2%; Score 1331; DB 25; Length 1452;  
Best Local Similarity 99.4%; Pred. No. 1.7e-294;  
Matches 1367; Conservative 0; Mismatches 5; Indels 3; Gaps 3;  
  
QY 219 GCGCCCCGGGATCGCTGGAGCCCGGATGCCACCA-GGGGCGCCCATGGGCCCCCC 277  
DB 78 GCGCCCCGGGATCGCTGGAGCCCGGATGCCACCAAGGGGGCGCCCATGGGCCCCCC 137  
  
QY 278 GGGCTCCCGGTACATGGGAGCCCGCGTGGACCCCGGCTGGCCCCCGGG-ATGGA 336  
DB 138 GGGCTCCCGGTACATGGGAGCCCGCGTGGACCCCGGCTGGCCCCCGGGCATGGA 197  
  
QY 337 GCGCGCCGCAAGCAGAGAGCCCGCCCGCC-GGCAGAGCCAGGCACAGAGCCAGGGCCA 395  
DB 198 GCGCGCCGCAAGCAGAGAGCCCGCCCGCCCGGGCAGAGCCAGGCACAGAGCCAGGGCCA 257  
  
QY 396 GCGGAGCCCAACCGCCCGCGAGCGCGAGTCCCAAGAGGAGGAGATGGCTGACAA 455  
DB 258 GCGGAGCCCAACCGCCCGCGAGCGCGAGTCCCAAGAGGAGGAGATGGCTGACAA 317  
  
QY 456 AATCTCTCCCTCAAAGGATTCGGGAGCTGGTCCCGAGTCCCGAGCTTACATGGACCTCTT 515  
DB 318 AATCTCTCCCTCAAAGGATTCGGGAGCTGGTCCCGAGTCCCGAGCTTACATGGACCTCTT 377  
  
QY 516 GCATTTGAGAGGAACTGGATCAAAACATCATATGCGGAAGGGGTGGACATCCAGGAGGC 575  
DB 378 GCATTTGAGAGGAACTGGATCAAAACATCATATGCGGAAGGGGTGGACATCCAGGAGGC 437  
  
QY 576 TCTGAAGAGGCCCATGAGCAAAAGCGGAAGCTGACTCTATATCTCAACACTTTTAA 635  
DB 438 TCTGAAGAGGCCCATGAGCAAAAGCGGAAGCTGACTCTATATCTCAACACTTTTAA 497  
  
QY 636 CCTCGGAAGTCTGATGTGAGGATTCGACCGGACGATTCGCTCTGGGAGCTACGGGT 695  
DB 498 CCTCGGAAGTCTGATGTGAGGATTCGACCGGACGATTCGCTCTGGGAGCTACGGGT 557  
  
QY 696 GGAGGGGAGCTCTGGATGATCCAGCAAAAGCGGAAGTCTCTCTTTCTTTTCAA 755  
DB 755 GGAGGGGAGCTCTGGATGATCCAGCAAAAGCGGAAGTCTCTCTTTCTTTTCAA 755

DB 558 GGAGGGGAGCTCTGGATGATCCAGCAAAAGCGGAGTCTCTCTTTCTTTCAA 617  
QY 756 GAGTTTGGTTCATCGAGCTGGACAAAGATCTTTATGGCCCTGACAAACACCTCTGAGTG 815  
DB 618 GAGTTTGGTTCATCGAGCTGGACAAAGATCTTTATGGCCCTGACAAACACCTCTGAGTG 677  
QY 816 GCATCGGACACCCACGACCCAGGAGAGCGGCTTCAGGTGAAACGGCCCTGGGACCT 875  
DB 678 GCATCGGACACCCACGACCCAGGAGAGCGGCTTCAGGTGAAACGGCCCTGGGACCT 737  
QY 876 GAGTGTGGTTCATCGAGCTGCTCTCATGCTGGACTACAGGCTCCCAAGTTCAAACCTGGA 935  
DB 738 GAGTGTGGTTCATCGAGCTGCTCTCATGCTGGACTACAGGCTCCCAAGTTCAAACCTGGA 797  
QY 936 TCCCCGGCTAGCCCGGCTGCTGGGCTGCACACAGAGCGCGCTCAGCAATTGTCCAGGC 995  
DB 798 TCCCCGGCTAGCCCGGCTGCTGGGCTGCACACAGAGCGCGCTCAGCAATTGTCCAGGC 857  
QY 996 CCTGTGGAGTATGTGAAGACCAACAGGCTGCAGGACTCCCATGACAAAGGAATACATCAA 1055  
DB 858 CCTGTGGAGTATGTGAAGACCAACAGGCTGCAGGACTCCCATGACAAAGGAATACATCAA 917  
QY 1056 TGGGGAAGTATTTCCAGCAGATTTTGAATTGCCCGGCTGAAGTTTCTGAGATTCC 1115  
DB 918 TGGGGAAGTATTTCCAGCAGATTTTGAATTGCCCGGCTGAAGTTTCTGAGATTCC 977  
QY 1116 CCAGCGCTCACAGCCCTGCTTATTTGCCCTGACCCCAATTGTCATCAACCATGTCATCAG 1175  
DB 978 CCAGCGCTCACAGCCCTGCTTATTTGCCCTGACCCCAATTGTCATCAACCATGTCATCAG 1037  
QY 1176 CGTGGAGCCCTTCAGACAGCAAGAGAGAGCGGTGCTATGACATTCAGTGGAGGTGGAGGA 1235  
DB 1038 CGTGGAGCCCTTCAGACAGCAAGAGAGCGGTGCTATGACATTCAGTGGAGGTGGAGGA 1097  
QY 1236 GCCATTAAAGGGGAGATGAGCAGCTTCTCTCTATCCAGCGCCCAACAGCAGGAGATCAG 1295  
DB 1098 GCCATTAAAGGGGAGATGAGCAGCTTCTCTCTATCCAGCGCCCAACAGCAGGAGATCAG 1157  
QY 1296 TCCCTCTGAGACAGTAAAGATCCATGAGAGAGATTTGAGTCCATAAACAGAGCTCAAGATCCAGAG 1355  
DB 1158 TGCTCTGACAGTAAAGATCCATGAGAGAGATTTGAGTCCATAAACAGAGCTCAAGATCCAGAG 1217  
QY 1356 GGAATTATGCTTAAGCTTCTCCAGAGACCCCAAGGCTATGTCCAAAGACCTGCTCCGCTC 1415  
DB 1218 GGAATTATGCTTAAGCTTCTCCAGAGACCCCAAGGCTATGTCCAAAGACCTGCTCCGCTC 1277  
QY 1416 CCAGAGCCGGAGCCTCAAGGTGATGACAGATGTAGCCGCAACCTGAGAGGAGCGCG 1475  
DB 1278 CCAGAGCCGGAGCCTCAAGGTGATGACAGATGTAGCCGCAACCTGAGAGGAGCGCG 1337  
QY 1476 GGCTGAGTTCTACCAACAGCCCTGGTCCAGGAGCGCTCAGTCCGCTACTTCTACTGCAA 1535  
DB 1338 GGCTGAGTTCTACCAACAGCCCTGGTCCAGGAGCGCTCAGTCCGCTACTTCTACTGCAA 1397  
QY 1536 GATCCAGCAGCGCAGGAGGAGCTGGAGCAGTCCGCTGTTGTGGCAACCTAG 1590  
DB 1398 GATCCAGCAGCGCAGGAGGAGCTGGAGCAGTCCGCTGTTGTGGCAACCTAG 1452

RESULT 2  
AAH98221  
ID AAH98221 standard; cDNA; 3089 BP.  
XX  
AC AAH98221;  
XT 12-OCT-2001 (first entry)  
XX Human EST-derived coding sequence SEQ ID NO: 78.  
DE Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
KW diagnostics; forensic test; gene mapping; genetic disorder;  
KW biodiversity; gene therapy; nutrition; ss.

XX	Homo sapiens.
OS	WO200154477-A2.
PN	02-AUG-2001.
PD	25-JAN-2001; 2001WO-US02687.
PF	25-JAN-2000; 2000US-0491404.
PR	17-JUL-2000; 2000US-0617746.
PR	03-AUG-2000; 2000US-0631451.
PR	15-SEP-2000; 2000US-0663870.
PA	(HYSE-) HYSEQ INC.
XX	Tang YF, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
P1	Cao Y, Dirmann KA, Zhang J, Weirman T;
DR	WPI; 2001-476164/51.
DR	P-PsDB; AAM23562.
XX	
PT	Isolated polypeptide for treatment of diseases, diagnostics, raising
FT	antibodies and research use -
PS	Claim 1; Page 231-232; 1275pp; English.
CC	The present invention provides the protein and coding sequences of novel
CC	proteins from a variety of organisms, including human, dog, cat, horse,
CC	cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC	urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC	from the organism of interest. They can be used in diagnostics,
CC	forensics, gene mapping, identification of mutations, to assess
CC	biodiversity and for nutritional purposes. The present sequence is a cDNA
CC	of the invention.
XX	
SQ	Sequence 3089 BP; 648 A; 954 C; 946 G; 541 T; 0 other;
Query Match	76.7%; Score 1322.6; DB 22; Length 3089;
Best Local Similarity	91.8%; Pred. No. 1.8e-292;
Matches 1637; Conservative.	0; Mismatches 79; Indels 67; Gaps 20
OY	1 GAATTCGGCGCAGGCGGCCGACGCGAGCGGAGCGAGCGGCGCGCTCCGG 60
DB	1 GAATTCGGCGCAGGCGGCCGACGCGAGCGGAGCGGAGCGGCGCGCTCCGG 60
OY	61 GCCGGGGTCCCGGGGAGCAGATTCTCAGAAATGCCCTTGTTGCAGCGCGGTGGC 120
DB	61 GCCGGGGTCCCGGGGAGCAGATTCTCAGAAATGCCCTTGTTGCAGCGCGGTGGC 120
OY	121 TCCGGGGCCAGGCAACCAGAGGGGGCACTGATGACTTCCAGGTGCAGAACCTTGCATCT 180
DB	121 TCCGGGGCCAGGCAACCAGAGGGGGCACTGATGACTTCCAGGTGCAGAACCTTGCATCT 180
OY	181 ATGATCTCCAGGTCTTTCAGACCCACCCACCGTAGTACAGCCGCCCGGATGCGTGTGA 240
DB	181 ATGATCTCCAGGTCTTTCAGACCCACCCACCGTAGTACAGCCGCCCGGATGCGTGTGA 240
OY	241 GCCCGGATGCCCCACCA-GGGGCGCCCATGGCCCCCGGGCTCCCGTACATGGGCAGC 299
DB	241 GCCCGGATGCCCCACCA-GGGGCGCCCATGGCCCCCGGGCTCCCGTACATGGGCAGC 300
OY	300 CCCGCGCTGCGAACCAGGCTGGGCCCCCGGGG-ATTGAAACCCGCCCGCAAGCAGTACG 358
DB	301 CCCGCGCTGCGAACCAGGCTGGGCCCCCGGGGATGAAACCCGCCCGCAAGCAGTACG 360
OY	359 CCCCGGCC-GGCAGAAGCAGGACAAGACCAAGGGGCAGCCGGAAGCCACCGCCCCGGC 417
DB	361 CCCCGGCCGAGGACCAAGGACAAGACCAAGGGCCAGTGGTCCACCGCCCCGGC 420
OY	418 CGGAGCCGCAATGCCAAGG-----GAGGAAGATGGCTGACAA 455
DB	421 CGGAGCCGCAATGCCAAGG-----GAGGAAGATGGCTGACAA 480

QY	466	AATCTTCCCTCAAAAGATTCGGGAGCTGTGTCCCGAGTCCAGGCTTACATGACCTTCT	515
Db	481	GAACCGCAGGAGCAGATTCGGGAGCTGTGTCCCGAGTCCAGGCTTACATGACCTTCT	540
QY	516	GGCATTTTGAAGAGAACTGGATCAACCATCATCGGAAGGGGGTGGACATCCAGAGGC	575
Db	541	GGCATTTTGAAGAGAACTGGATCAACCATCATCGGAAGGGGGTGGACATCCAGAGGC	600
QY	576	TCTGAAGAGGCCCATGAAACAAAACCGAGACTGCGACTTATATCTTCCAACTTTTAA	635
Db	601	TCTGAAGAGGCCCATGAAACAAAACCGAGACTGCGACTTATATCTTCCAACTTTTAA	660
QY	636	CCCTGCGCAAGTCTGATCTGAGGATTTCCGAGCGCAGACTTCCCTCTGGAGCTAACGGT	695
Db	661	CCCTGCGCAAGGCTGATCTGAGGATTTCCGAGCGCAGACTTCCCTCTGGAGCTAACGGT	720
QY	696	GGAGGGGAGGCTCCGAGTGTAT-----CCGAGCAAAACAGAAAGCGGAAAT	739
Db	721	GGAGGGGAGGCTCCGAGTGTATGATGATCGTCCGGGCCAGGCCAGCAAAACAGAGCGGAAGT	780
QY	740	TCTCTTCTTTCTTCAAGATTTGGTCATCGAGCTGACAAAGATCTTTATGGCCCTGACA	799
Db	781	TCTCTTCTTTCTTCAAGATTTGGTCATCGAGCTGACAAAGATCTTTATGGCCCTGACA	840
QY	800	ACCACCTGTTGATGATGCGATCGGACACCCACGACCCGAGAGACGAGCGGCTTCCAGGTGA	859
Db	841	ACCACCTGTTGATGATGCGATCGGACACCCACGACCCGAGAGACGAGCGGCTTCCAGGTGA	900
QY	860	AACGGCCTGGGGGACCTGATGTGGCGCTGCACGCTGCTCTCATCTGAGACTACAGCCTC	919
Db	901	AACGGCCTGGGGGACCTGATGTGGCGCTGCACGCTGCTCTCATCTGAGACTACAGCCTC	960
QY	920	CCCAATTCAAACTGGATCCCGCGCTAGCCCGGCTGTGGGGCTGCACACAGAGCCGCT	979
Db	961	CCCAATTCAAACTGGATCCCGCGCTAGCCCGGCTGTGGGGCTGCACACAGAGCCGCT	1020
QY	980	CAGCATTTGTCCAGGCCCTGTGGGAGATATGTGAAGACCAACAGGCTGAGGACTTCCATG	1039
Db	1021	CAGCATTTGTCCAGGCCCTGTGGGAGATATGTGAAGACCAACAGGCTGAGGACTTCCATG	1080
QY	1040	ACAAGGAATACATCATGATGAGGGAGCAAGATTTTCCAGCAGATTTTGTATGTGTCGGGCTGA	1099
Db	1081	ACAAGGAATACATCATGATGAGGGAGCAAGATTTTCCAGCAGATTTTGTATGTGTCGGGCTGA	1140
QY	1100	AGTTTCTGAGATTTCCCGAGCGCTCAACGCGCTGTATTTGCCCTTGACCCCAATTGTCA	1159
Db	1141	AGTTTCTGAGATTTCCCGAGCGCTCAACGCGCTGTATTTGCCCTTGACCCCAATTGTCA	1200
QY	1160	TCAACCATGTATCATGCGTGGACCCCTTACAG-CCAGAAGAAAGCAG-CGTGCTATGACAT	1217
Db	1201	TCAACCATGTATCATGCGTGGACCCCTTACAG-CCAGAAGAAAGCAGGTGTGCTATGACAT	1260
QY	1218	TGACGTGAGGTGGAGAGGCC--ATTAAAGGGGAGATGAGCAGCTTCTCT--CCTATCCA	1273
Db	1261	TGACGTGAGGTGGAGAGGCCCATTTAAAGGGGCGAGATGAGCAGCTTCTCTCTATTTCCA	1320
QY	1274	CGGCCAACCAAGCAGAGATCAGT---CCTCTGACAGTAAGAT-CCATGAGCAGATTGA	1328
Db	1321	CGGCCAACCAAGCAGAGAGATTCAGTGTCTTGGACAGTAAGATCCCATGAGCGGATTTGA	1380
QY	1329	GTCCTATAACCAAGCTCAAG----TCCAGAGGAGCTTC-ATGCTTAAGCTTCCAGAGA-1382	
Db	1381	GTCCTATAACCAAGCTTCAGAGATCCAGAGGGACTTCAATGCTTAAGATTTCTTCCAGAG-1440	
QY	1383	--CCCAAGGCTATGTTCACAAGCCTGTCCGCTCCGATCCAGAGCCGGGACTTCAAGG--TGA	1438
Db	1441	ACCCCAAGGCTATGTTCACAAGCCTGTCCGCTCCGATCCAGAGCCGGGACTTCAAGGTTGA	1500
QY	1439	TGACAGATGTAGCCGGCAACCTTGAAGAGAGGCGCCGGGC-TGATTTCAACACC-AGCC	1496
Db	1501	TGACAGATGTAGCCGGCAACCTTGAAGAGAGGCGCCGGGCTTGAATTTCAACCAAGGCC	1560





```

Db      1081 ACAAGAAATACATCATATGGGACAGATATTTCCAGAGATTTTGTATGTCTCCCGGCTGA 1140
QY      1100 AGTTTCTAGATATCCCGAGCGGCTCAGACCGCTGTATTTGGCCCCCGTACCATATGTGA 1159
Db      1141 AGTTTCTAGATATCCCGAGCGGCTCAGACCGCTGTATTTGGCCCCCGTACCATATGTGA 1200
QY      1160 TCACCATGTATCATAGCGTGAACCTTTCAGA--CCAGAGAAGACAG--CGTGTATGACAT 1217
Db      1201 TCACCATGTATCATAGCGTGAACCTTTCAGA--CCAGAGAAGACAGCGTGTATGACAT 1260
QY      1218 TGACGTGAGGTGAGAGAGCC--ATTAAAGGGCAGATGAGCCTTCT--CCTATCCA 1273
Db      1261 TGACGTGAGGTGAGAGAGCCCATTAAGGGCAGATGAGCCTTCTCTTCTATTTCA 1320
QY      1274 CGGCCACAGCAGAGATCAGT---CCTTGGAGATGTAAT--CCATGAGACATTGA 1328
Db      1321 CGGCCAAACAGCAGAGATCAGTCTTCTGGAGATGTAATCCATGAGCCGATTTGA 1380
QY      1329 GTCCATTAACAGCTCAGA---TCCAGAGGGAGCTTC--ATGCTAGCTTCTCCAGAGA- 1382
Db      1381 GTCCATTAACAGCTCAGATCCAGAGATCCAGAGGAGCTTCAATCTTAAGTTTCTTCCAGAG 1440
QY      1383 --CCCAAGGCTATGTCCAGACCTGCTCCGCTCCAGAGCGGGAGCCTCAAG--TGA 1438
Db      1441 ACCCCAAAGGCTATGTCCAGACCTGCTCCGCTCCAGAGCGGGAGCCTCAAGTTGA 1500
QY      1439 TCACGATGTAGCCGCAACCTTGAAGAGAGAGCGCGGCG--TGATTTTCAACC--AGCC 1496
Db      1501 TCACGATGTAGCCGCAACCTTGAAGAGAGAGCGCGGCGCTTGATTTTCAACCAGCC 1560
QY      1497 CTGGTCCAGAGAGCGGCTCAGTC--GGTACTTTAC--TGCAAGATCCAGAGAGCGAGGAG 1554
Db      1561 CTGGTCCAGAGAGCGGCTCAGTCCTGCTACTTCTTCTGAGATCCAGAGAGCGAGGAG 1620
QY      1555 GAGCTGAGCAGTCTGCTGTTGTGCCAACACCTTAGAGAGCCCAAAAACAGAGCAGCAG 1614
Db      1621 GAGCTGAGCAGTCTGCTGTTGTGCCAACACCTTAGAGAGCCCAAAAACAGAGCAGCAG 1680
QY      1615 GGAATCTTTCAGCCGTGTCCCGGGCCCGCAGCATTTTCCCGGGCTCCAGC--TCACTCTTC 1673
Db      1681 GGAATCTTTCAGCCGTGTCCCGGGCCCGCAGCATTTTCCCGGGCTCCAGCATCCTCTC 1740
QY      1674 TCACACCTTGGGGTGTGGGGCTGGATTAAAGCATTCATCTG 1716
Db      1741 TCACACCTTGGGGTGTGGGGCTGGATTAAAGCATTCATCTG 1783

RESULT 4
AB054477
ID      AB054477 standard; cDNA; 1970 BP.
XX
XX
AC      AB054477;
XX
XX
DT      22-AUG-2002 (first entry)
XX
XX
DE      Human ovarian antigen HSD108 cDNA, SEQ ID NO:357.
XX
XX
KW      Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW      ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW      infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW      PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW      inflammatory condition; immune disorder; blood disorder;
KW      cardiovascular disorder; respiratory disorder; neurological disorder;
KW      gastrointestinal disorder; urinary system disorder; drug screening;
KW      gene therapy; chromosome mapping; forensic analysis;
KW      antibody preparation; cytotoxic; immunomodulatory; neuroprotective;
KW      antiinflammatory; gynaecological; reproductive; gene; ss.
XX
OS      Homo sapiens.
XX
XX      WO200200677-A1.
XX

```

```

PD      03-JAN-2002.
XX
XX      07-JUN-2001; 2001WO-US18569.
XX
XX      07-JUN-2000; 2000US-209467P.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX
XX      Birse CE, Rosen CA;
XX
XX      WPI; 2002-147878/19.
XX
XX      P-PSDB; ABP41400.
XX
XX      Isolated nucleic acid molecules encoding novel ovarian polypeptides,
XX      PT      useful in the prevention, treatment and diagnosis of cancer (e.g.
XX      PT      ovarian cancer), immune disorders, cardiovascular disorders and
XX      PT      neurological diseases -
XX
XX      Claim 1; SEQ ID No 357; 2922bp; English.
XX
XX      The invention relates to 2175 novel human ovarian antigens (ABP41054-
XX      CC      ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
XX      CC      encompasses polypeptides 90% identical and polynucleotides 95% identical
XX      CC      to the sequences of the invention. The invention additionally relates to
XX      CC      recombinant vectors and host cells comprising human ovarian antigen
XX      CC      polynucleotides, antibodies against human ovarian antigens, and the use
XX      CC      of ovarian antigen polynucleotides and polypeptides in diagnosing,
XX      CC      treating, prognosting or preventing various ovary and/or breast-related
XX      CC      disorders. Such conditions include ovarian cancer and breast cancer, and
XX      CC      metastatic tumours of ovarian or breast origin, reproductive system
XX      CC      disorders (e.g., infertility, disorders of pregnancy, anovulation,
XX      CC      polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
XX      CC      disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
XX      CC      shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
XX      CC      vaginitis), immune disorders (e.g., congenital and acquired
XX      CC      immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
XX      CC      blood-related disorders (e.g., anaemia), cardiovascular disorders,
XX      CC      respiratory disorders, neurological disorders, gastrointestinal disorders
XX      CC      and urinary system disorders. Ovarian antigen polypeptides and
XX      CC      polynucleotides may also be used in screening for compounds which
XX      CC      modulate ovarian antigen expression or activity. The polynucleotides may
XX      CC      further be used for gene therapy, chromosome mapping, in the
XX      CC      identification of individuals and in forensic analysis, and the
XX      CC      polypeptides may be used as food additives or to prepare antibodies
XX      CC      useful in disease diagnosis, drug targeting and phenotyping. The present
XX      CC      sequence represents cDNA encoding a human ovarian antigen of the
XX      CC      invention.
XX      CC      Note: The sequence data for this patent did not form part of the printed
XX      CC      specification, but was obtained in electronic format directly from WIPO
XX      CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 1970 BP; 463 A; 590 C; 522 G; 389 T; 6 other;
XX
Query Match      57.0%; Score 983; DB 24; Length 1970;
Best Local Similarity 99.0%; Pred. No. 6; 7e-215;
Matches 1027; Conservative 3; Mismatches 3; Indels 4; Gaps 4;
QY      46 GCGGCGGCTCCGGGCGGCGGTCCGGGGAGCAGATCTTCAGATGGCCCTTGCT 105
Db      1 GCGGCGGCTCCGGGCGGCGGTCCGGGGAGCAGATCTTCAGATGGCCCTTGCT 60
QY      106 GAGGCGGCGGCGGTCCGGGCGGCGGAGCAGAGGGGCGACGTGACTCTCCAGGTCC 165
Db      61 GAGGCGGCGGCGGTCCGGGCGGCGGAGCAGAGGGGCGACGTGACTCTCCAGGTCC 120
QY      166 AGGACCTTGCCATCTATGACTTCAGAGTTCAGACCCAGCCAGCGGTACAGCGCCC 225
Db      121 AGGACCTTGCCATCTATGACTTCAGAGTTCAGACCCAGCCAGCGGTACAGCGCCC 180
QY      226 GGGATCCGCTGAGCGCCGAGTGCACCA--GGGCGGCCCATGAGCCCTCCGCGCTCC 284
Db      181 GGGATCCGCTGAGCGCCGAGTGCACCA--GGGCGGCCCATGAGCCCTCCGCGCTCC 240

```





QY 735 GAAGTTCCTCTTCTTCAAGAGTTTGGTCAATCGAGCTGGACAAGATCTTTATGGCCC 794  
Db 684 AAGTTCCTCTTCTTCAAGAGTTTGGTCAATCGAGCTGGACAAGATCTTTATGGCCC 743  
QY 795 TCACAACACCTCGTTGAGTGCATCGACACCCACGACCCAGGAGAGCGGCTTCCA 854  
Db 744 AGACAACCATCTGGTAGAATGCAAGGACCGCCACTACCCAGGAGACGATGGCTTCCA 803  
QY 855 GGTGAAACGGCTCGGGACCTCAGTGTGGCTGCACGCTGCTCTCATCTGCTGGACTACA 914  
Db 804 GGTGAAGCGGCCAGAGATGCTGAATGACGGTGTACTGCTGCTGATGCTGGACTACA 863  
QY 915 GCTTCCCACTTCAAACTCGATCCCGCTAGCCCGCTGCTGGGGCTGCACACAGAG 974  
Db 864 GCGCCCGCAGTTTAAATTAGACCTCGCTCGCTCGCTCTTGGGCATCCATACCCAGAC 923  
QY 975 CCGCTCAGCCATGTCAGGCGCTGTGGCAGTATGTGAAGCAACACAGCTGCAGGACTC 1034  
Db 924 ACGTCCAGTGATCATCAAGCACTGTGCAGTATATTAAACACACAAGCTCCAGGACCC 983  
QY 1035 CCATGACAAGGAATACATCAATGGGGACAAGTATTTCCAGCAGATTTTGTGATTTGCCCG 1094  
Db 984 TCACGAGCGAGATTTGTTCTCTGTGACAGTACCTCCAGCAGATCTTTGATCTCAGG 1043  
QY 1095 GCTGAAGTTTCTGAGATTCCTCAGCGCTCTCAGCGCTGCTATTGCCCCCTGACCCAAAT 1154  
Db 1044 GATGAAGTTCTCAGAGATCCCTCAGCGCTCTCAGCGCTGCTATTGCCCCCTGACCCAAAT 1103  
QY 1155 TGTCAATCAACATGTCATCAGCGTGGACCCCTTCAGACCAAGAGACAGCGTGTATGA 1214  
Db 1104 CATCATCAATCATGTTCATCAGTGTGGACCCCAATGACCAGAAAAAGACCGGTGTATGA 1163  
QY 1215 CATTGACGTGGAGTGGAGGCCATTAAAGGGGAGATGAGCAGCTTCTCTCTATCCAC 1274  
Db 1164 CATTGACGTGGAGTGGAGTGACACTCTGAAGCCACAGATGAATCTTTCTCTGTGTCAC 1223  
QY 1275 GCGCAACACGACGAGATCAGTCCTCTGACAGTAAAGATCCATGAGACGATGAGTCCAT 1334  
Db 1224 TGCCAGCCAGCAGGAGATCGCCACTCTAGACAAAGATCCATGAGCAGATAGAGACCAT 1283  
QY 1335 AAACAGCTCAAGATCCAGAGGACTTCATGCTAAGCTTCTCCAGAGACCCCAAGGCTA 1394  
Db 1284 CAACAGCTGAAGACCCAGCAGAGTTTCATGTTGAGCTTGCCC---GAGCCCTCAGGTTT 1340  
QY 1395 TGTCGAAGACCTGCTCCGCTCCAGAGCGGACCTCAAGGTGATGACAGATGTAGCCGG 1454  
Db 1341 CATCAATGATGGCTTCAGTCCAGTGCAGGGGCTTCAGACGATGACTGATGTTGGG 1400  
QY 1455 CAACCTGAAGAGAGCGCGGCTGAGTTCTACCAAGCCCTGGTCCAGGAGCGGT 1514  
Db 1401 TAAACCGGAAGAGAGCGTCTGCTGAGTTCTTACTTCCAGCCCTGGGCTCAGGAGGCTGT 1460  
QY 1515 CAGTGCCTTCTTACTGCAAGATCCAGAGCGCAGGAGCTGGAGCAGTCCCTGGT 1574  
Db 1461 GTGCCGATATCTTACTCCAGGTCGACGAGAGCGGCAAGAGTTAGAGAACGCCCTGG 1520  
QY 1575 TGTGCGCAACACTAGGAGCCC 1596  
Db 1521 AATCCGAACACATAGAGGCGCTC 1542

RESULT 7  
ABS73872  
ID ABS73872 standard; cDNA; 2511 BP.  
XX AC ABS73872;  
XX DT 06-DEC-2002 (first entry)  
XX DE Human cDNA encoding NAAP12, Incyte 3112390Cb1.  
XX Human; ss; gene; nucleic acid-associated protein; NAAP; gene therapy;  
KW microarray; proliferative disorder; developmental disorder;

cardiovascular disorder; neurological disorder; autoimmune disorder; inflammatory disorder; atherosclerosis; bursitis; cirrhosis; hepatitis; psoriasis; cancer; lymphoma; melanoma; brain cancer; breast cancer; Cushing's syndrome; Alzheimer's disease; AIDS; Creutzfeldt-Jakob disease; acquired immunodeficiency syndrome; Goodpasture's syndrome; Crohn's disease; multiple sclerosis.  
Homo sapiens.  
WO200274913-A2.  
26-SEP-2002.  
14-MAR-2002; 2002WO-US07869.  
16-MAR-2001; 2001US-276857P.  
19-APR-2001; 2001US-285489P.  
19-APR-2001; 2001US-285558P.  
04-MAY-2001; 2001US-288646P.  
04-MAY-2001; 2001US-288700P.  
10-MAY-2001; 2001US-290510P.  
11-MAY-2001; 2001US-290369P.  
16-NOV-2001; 2001US-332426P.  
(INCY-) INCYTE GENOMICS INC.  
Hillman JL, Baughn MR, Swarnakar A, Yue H, Elliott VS, Burford N; Ding L, Tang YT, Lee SY, Azimzai Y, Wallia NK, Gietzen KJ; Griffin JA, Lai PG, Yang J, Borowsky ML, Richardson TW, Yue H; Becha S, Forsythe IJ, Jones KA, Warren BA, Thangavelu K; Honchell CD, Jolley HE, Hafalia AJA, Ring HZ; WPI; 2002-723448/78.  
P-PSDB; ABG95655.  
New human nucleic acid-associated proteins and polynucleotides, useful for diagnosing, treating or preventing cardiovascular, neurological, autoimmune or inflammatory disorders, e.g. atherosclerosis, Alzheimer's or AIDS -  
Claim 5; Page 205-206; 219pp; English.  
The invention relates to an isolated human nucleic acid-associated protein (NAAP), comprising the protein sequences appearing as ABG95644-ABG95669, or an immunogenic or biologically active fragment. Also included are the NAAP encoding polynucleotides, a recombinant polynucleotide comprising a promoter sequence operably linked to the NAAP polynucleotide, a cell transformed with the recombinant NAAP polynucleotide, an NAAP transgenic organism, an anti-NAAP antibody, screening compounds as modulators or ant/agonists of NAAP or modulators of NAAP polynucleotide expression, a microarray comprising 30-60 nucleotides of the NAAP polynucleotides and generating an expression profile of a sample that contains NAAP polynucleotides. The polypeptide, polynucleotide and agonist are useful for treating a disease or condition associated with decreased expression of functional NAAP. The antagonist is useful for treating a disease or condition associated with overexpression of functional NAAP. The antibody that specifically binds to the polypeptide is useful for diagnosing a condition or disease associated with the expression of NAAP. These diseases include proliferative, development, cardiovascular, neurological, or autoimmune or inflammatory disorders. These polypeptides, polynucleotides, agonists and antagonists may also be used for preventing these diseases. These disease or conditions associated with functional NAAP expression also includes atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis, cancers (e.g. lymphoma, melanoma, brain cancer or breast cancer), Cushing's syndrome, Alzheimer's disease, acquired immunodeficiency syndrome (AIDS), Creutzfeldt-Jakob disease, Goodpasture's syndrome, Crohn's disease, or multiple sclerosis and many other diseases and conditions listed in the specification. The present sequence encodes an NAAP protein.  
Sequence 2511 BP; 512 A; 713 C; 760 G; 526 T; 0 other;

Query Match 37.2%; Score 642; DB 24; Length 2511;  
 Best Local Similarity 71.1%; Pred. No. 6.4e-137;  
 Matches 880; Conservative 0; Mismatches 325; Indels 33; Gaps 1;

QY 393 CCAGCCGAGCCACCGCCCGCGGAGCCGCAATGCCAAGAGAGAGAGATGGCTTA 452  
 DB 445 CCAGGCGGAGCCCTCCATGCTGCTCCAGCGCGGGGTTAAAGAGAGAGATGGCAGA 504  
 QY 453 CAAATCCCTCCCAAGAGATTCGGAGGCTGCCGAGTCCAGGCTTACATGGAGCT 512  
 DB 505 TAAAGTTTACCTCAGCGAATCGGAGGCTTGTTCAGAGTCTCAGGCGTACATGGATCT 564  
 QY 513 CTTCGATTTGAGAGAGAACTGATCAACATCATCGGAAAGGGGGTGAATCCAGGA 572  
 DB 565 CTTCGCTTTGAGCGGAACTGGACCAAGACATTCGTCGCCAGCGGATGGAGATCCAGA 624  
 QY 573 GGCTCTGAAGAGCCCATGAAAGCAAAAGCGGAAGCTGCGAATCTATATCTCCACA 632  
 DB 625 GGCATCAAAAAGCTCTGACACAAAAGCGAAAGTTCCGATCTCATTTCCAAATCGTT 684  
 QY 633 TAACTCTGGAACTGTGATGCTGAGAT----- 660  
 DB 685 CAGTCCAGCAAGGCGGAAAGCGCATAGTACAGAACTGACAGGAGCCCTGGGGAGCC 744  
 QY 661 -TCCGAGCGCAGCATTCCTCTGGAGCTACCGGCTGAGAGGGAAGCTCCTGGATATCC 719  
 DB 745 AGCAGGGGAGACAGTGGCTTCTGGGAATCCGAGTGGAAAGAAAAGCTGCTGATATCC 804  
 QY 720 CAGCAAAAGAAAGCGGAAGTCTCTTCTTCTTCAAGATTTGGTATCGAGCTGACAA 779  
 DB 805 TAGCAAAAGAAAGAGAGTTTCTTCTTCTTAAAGAGCTGCTGATGAGCTGACAA 864  
 QY 780 AGATCTTTATGCGCTTCGACACCACTCTGTAAGTGCATCGGACCCACCAAGCA 839  
 DB 865 GAGAGCTGTACGGGCTGAAATCACTGTGTGAGTGCACCGGATCCACCAAGCA 924  
 QY 840 GACGAGCGGCTTCAGGTGAAAGCGGCTGGGAGCTGAGTGTGCGCTGACGCTGCTCT 899  
 DB 925 GACAGATGCTTCCAGTAAAGCGGCTGAGAGCTTCAAGCTCAAGTGCACCTCTGCT 984  
 QY 900 CATGCTGACATCAAGCTTCCAGTTCAAACTGATTCGCCGCTTACCGGCTGCTGG 959  
 DB 985 CATGCTGATCATCAAGCTTCCAGTTCAAACTGATTCGCCGCTTACCGGCTGCTGG 1044  
 QY 960 GCTGCACACAGAGCGGCTCAGCCATTGTCAGGCGCTTGGCAATGTAAGCA 1019  
 DB 1045 AGTGCACACAGAGCGGCTCAGCCATTGTCAGGCGCTTGGCAATGTAAGCA 1104  
 QY 1020 CAGGCTGACAGACTCCCATGACAAAGATATCATGAGGGAACAATTTCCAGCAGAT 1079  
 DB 1105 CAGCTGACAGATGGGACAGGCGGAGTACATCACTGACCGTACTTCCGCGAGAT 1164  
 QY 1080 TTTTGTATTCCTCCGCTGAAGTTTCTGAGATTTCCCAAGCGCTCAGACCTGCTATT 1139  
 DB 1165 CTTCATTTGTGGCGACTCCGTTTCTCCGAGATTTCCCAAGAGCTGGCAGGTTGCTGA 1224  
 QY 1140 GCGCCCTGACCAATTTGATCAACCATGTATCAGCTGAGACCTTTCAGACCAAGAA 1199  
 DB 1225 GATCCAGAGCCCATGTTGATCAACCATGTATTAAGTGTGACCTTAAGCAGAGAA 1284  
 QY 1200 GACAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1259  
 DB 1285 GACAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1344  
 QY 1260 CTTCCTCTATCCAGCGGCAACAGCAGAGAGATCATCTCTGAGCAGTAAAGTCCATGA 1319  
 DB 1345 TTTTGTGGGCTCTACACCAATCAGAGAGAGATCGGCTCCTTGAATGATGATGATGAT 1404  
 QY 1320 GACGATTTAGTCCATTAACAGCTCAAGATTCAGAGGAGCTTATGCTTCTCCAG 1379  
 DB 1405 GACCATTTAGTCCATTAACAGCTCAAGATTCAGAGGAGCTTATGCTTCTCCAG 1464  
 QY 1380 AAGACCCCAAGAGTATGTCAAGACTGCTCCGCTCCAGAGCGGAGACTCAAGGTAT 1439

DB 1465 CACACCCGAGGACTTCATCCAGAGATGCTCGTTCCAGCGCCAGACCTCAAGATCAT 1524  
 QY 1440 GACCATATGACCGGCAACCTTGAAGAGAGCGCGGCTAGTTCACACACCCCTG 1499  
 DB 1525 CACTATGTATTTGAAATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1584  
 QY 1500 GTCCAGAGAGCGCGTCACTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1559  
 DB 1585 GCGCCAG 1644  
 QY 1560 GGAGCAGTCTGCTGTTGCGCAACACCTAGAGAGCA 1597  
 DB 1645 GGAACAGGTGCTGGGATTCGCTGACCTACTGCTCA 1682

RESULT 8  
 ABL02583  
 ID ABL02583 standard; cDNA; 2010 BP.  
 XX  
 AC ABL02583;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2231.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EM;  
 XX  
 DR WPI, 2001-656860/75.  
 XX  
 DR P-PSDB; ABB58480.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 2231; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (ABBS7737-ABP2072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 CC Sequence 2010 BP; 509 A; 550 C; 545 G; 406 T; 0 other;  
 SQ

Query Match 33.2%; Score 572.6; DB 23; Length 2010;  
 Best Local Similarity 69.9%; Pred. No. 4.4e-121;  
 Matches 824; Conservative 0; Mismatches 334; Indels 21; Gaps 3;

QY 433 AAGAGAGAGAGATGCTGACAAATCTCCCTCAAGATTCGGAGCTGCTCCCGAG 492

Db 581 AAGAAAAAGAGCTGGCGCAAAAGATCTGCGCGAAGAGGTGCGGGATCTCGTGGCCGAA 640  
QY TCCAGGCTTACATGCACTCTTGGCAATTTAGAGGAAACTGGATCAACCATCATGCGG 552  
Db 641 TCGCAAGCGTACATGGAATCTGCTGACGTTTGAGCGAANAAGTGGATGCCACCATATGCGC 700  
QY AAGGGGTGGACATCCAGAGGCTCTGAAGAGGCCCATGAAGCAAAAGCGGAAGCTGGCA 612  
Db 701 AAACGCTCGACATCCAGAGGCCCTCAAGGGCCCCCATGAAGCAAGCGCAAGCTGGCG 760  
QY CTCTATATCTCCACACTTTTAACCTCGAGTCTGATGCTGAGGATTCG---GACGGC 669  
Db 761 ATCTTCATCTCGAACACATTTATCCAGAGGAGCCCAAAATGATGGCGAGAGGGC 820  
QY AGCATTGCTCTCGGAGTACGGGTGGAGGGGAAGCTCTCGGATGATCCC----- 720  
Db 821 GCCGTGCTCTTGGGAATTCGGGTGGAGGTGCGCTTTAGAGGATGCGAAGGGCGAT 880  
QY 721 -----AGCAAAAGAGCGGAAGTTCTTCTTTCAAGAGTTTGGTCAATCGAGCTG 774  
Db 881 CCCAATACAAAGATCAAAACGCAAGTTTTCGTGTTTTCAAGTCGCTGTTATCGAGCTG 940  
QY 775 GACAAGATCTTATGGCCCTGACAAACCACTCTGTGAGTGCGCATCGCACACCCACGACC 834  
Db 941 GATAAGGAACTGTACGGTCCGAGCAACCATCTGGTTCGAGTGGCATCGCACTCAACACC 1000  
QY 835 CAGGAGCGGAGCGGCTTCCAGGTGAACGGCTCGGGACCTGAGTGTGCGCTGCACGCTG 894  
Db 1001 CAGGAGCGGAGCGGTTTCAGGTGAAGCGCGCGGCGATCGCATGTGCGCTGCACCATC 1060  
QY 895 CTCTCATCTGGAATACACAGCTCCCGAGTTCAAACCTGGATTCGCCGCTAGCCGGCTG 954  
Db 1061 CTCTGCTGCTGTACTACAGCGGCTGCAGTTCAGCTGGACCCCGACTCGCAGGTTG 1120  
QY 955 CTGGGCTGCACACAGAGCGCTCAGGCATTGTCCAGGCCCTGTGSCAGTATGTGAAG 1014  
Db 1121 CTGGCGGTACACACAGACAGGCGGCTGTATATATCCGCGCTGTGCGAGTACATCAAG 1180  
QY 1015 ACCAAGCGGTGAGAGCTCCCATCAAGGAATACATCAATGGGGCAAGTATTTCAG 1074  
Db 1181 AGGCAAGCTTACAGGATGCCCAAGCGGGAGTACATCAATTCGCAAGATATCTGGAG 1240  
QY 1075 CAGATTTTGTATGTTCCCGGCTGAAGTTTCTGAGATTCGCCAGCGCTCACAGCCCTG 1134  
Db 1241 CAGATATTTCAGTTGCGAGCGGATGAAGTTTGCAGATACCGCAACGCTCAATCCGCTG 1300  
QY 1135 CTATTGCCCCCTGACCCAAATGTCTATCAACATGTCTCAGCGTGGACCTTCAGACAG 1194  
Db 1301 CTGCATCCGCCCGATCCGATTTGTATCAATCATTTTCAT---TGAGAGCGGTGCAGAGAAC 1357  
QY 1195 AAGAGACAGCGTGTATGACATTCACGTTGAGGTGGAGGAGCCATTAAAGGGGCGAGATG 1254  
Db 1358 AAGCAGATGCTGCTAGACATCGATGTGAGGTGGATGATACGCTCAAAAACAGATG 1417  
QY 1255 AGCAGCTTCTCTTATCCACGCCAACACAGCAGGAGATCAGTCTCTCGGACAGTAAGATC 1314  
Db 1418 AACAGCTTCTCTGATGAGCACTCGAGCCACAGAGGATCCAGGGGCTGCACACCAAGATC 1477  
QY 1315 CATGAGACGATGAGTCCATAAACAAGCTCAAGATCCAGAGGAGTTCATGCTAAGCTTC 1374  
Db 1478 CACGAGCGGTGGACAGCATCAACAGATGAAGACGAAACAGGGGATTTCTTCTAAGCTTC 1537  
QY 1375 TCCAGAGACCCCAAGGCTATGTCCAAAGACTGCTCCGCTCCAGAGCGGAGCTCAAG 1434  
Db 1538 GCCAAGGATCCCAAAATGTTTATTCATCTGCTGGATCATAGTGACAGAGGATTTGAAG 1597  
QY 1435 GTGATGACAGATGTAGCCGCAACCTTGAAGAGGAGCGCGGCTGAGTTCTTACCAACAG 1494  
Db 1598 CTGATGACGATGTAGCTGGCAATCCGAGGAGGAGCGCTCGGGCGGAGTTCTTATTACCAG 1657  
QY 1495 CCCTGGTCCAGGAGCGCTAGTCTGCTACTTCTACTCAAGATCCAGAGCGGAGGAG 1554  
Db 1658 CCATGGACGACGAGGCGGCTTTCGCGCTACTTCTTTCACCAAGGTTCAACAGAGCGGCGC 1717

QY 1555 GAGCTGAGCAGTGGCTGGTTGTGCGCAACACCTAGGAG 1593  
ID ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1718 GAATTGGAGCAGCGCTGGGCATACGCAACGGCTAGGTG 1756

## RESULT 9

ABX34564  
ID ABX34564 standard; cDNA; 2581 BP.

AC ABX34564;

DT 13-FEB-2003 (first entry)

XX Human mdt cDNA SEQ ID 125.

XX MDDT; human; disease detection and treatment molecule polypeptide;  
KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;  
KW gene therapy; protein replacement therapy; cell proliferative disorder;  
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;  
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
KW psoriasis; hepatitis; gene; ss.

XX Homo sapiens.

XX OS

XX WO200279449-A2.

XX PN

XX FD 10-OCT-2002.

XX 27-MAR-2002; 2002WO-US09944.

XX 28-MAR-2001; 2001US-279619P.

XX 29-MAR-2001; 2001US-280067P.

XX 29-MAR-2001; 2001US-280068P.

XX 16-MAY-2001; 2001US-291280P.

XX 17-MAY-2001; 2001US-291829P.

XX 17-MAY-2001; 2001US-291849P.

XX 19-JUN-2001; 2001US-299428P.

XX 20-JUN-2001; 2001US-299776P.

XX 20-JUN-2001; 2001US-300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;

XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Klesfeld Y, Gerstein EH;

XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX WPI: 2003-058431/05.

XX P-PSDB; ABU11574.

XX New purified disease detection and treatment molecule proteins and

XX polynucleotides, useful for diagnosing, treating or preventing cancers

XX (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis

XX or hepatitis.

XX Claim 1; SEQ ID NO 125; 339pp + Sequence Listing; English.

XX This invention describes a novel disease detection and treatment molecule  
XX polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,  
XX osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,  
XX antianaemic, antipsoriatic and hepatotropic activity. The polynucleotides  
XX and the polypeptides of the invention can be used for gene therapy,  
XX protein replacement therapy and are useful for treating a variety of  
XX diseases or conditions. These polypeptides or polynucleotides are  
XX particularly useful for diagnosing, treating or preventing cell  
XX proliferative disorders (e.g. cancers including adenocarcinoma,  
XX leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
XX disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
XX syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or

CC hepatitis. ABX3440-ABX34835 encode the MDR1 polypeptides represented in CC ABU11450-ABU11845, described in the disclosure of the invention.  
CC NOTE: The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 2581 BP; 572 A; 652 C; 737 G; 620 T; 0 other;

Query Match 29.7%; Score 511.4; DB 25; Length 2581;

Best Local Similarity 72.1%; Pred. No. 4,6e-107;

Matches 690; Conservative 0; Mismatches 246; Indels 21; Gaps 1;

```

OY 662 CCGACGCGACGATTCCTCTGAGAGCTACGAGGTGAGGAGGAGCTCTGATGATCCCA 721
DB 773 CAGGGGACAGAGGCTTCTCTGGAATCTCGAGTGAAGAAAAGCTGAGTATCTTA 832
OY 722 GCAAAAGAGGAGGAGTCTCTTCTTCTTCAAGATTGATCATGAGCTGAGCAAG 781
DB 833 GCAAAAGAGGAGGAGTCTTCTTCAATTTTAAAGCCCTGATTAAGCTGAGCAAG 892
OY 782 ATCTTATGCGCTGCAACACCTCTGTG-----AGTGGCATC 820
DB 893 AGCTGTACCGGCTGCAATCACTGTGAGCATCTTCCCTCCCTGCAAGTGGCACC 952
OY 821 GGAACACCAAGACCGAGAGAGAGGAGGAGGAGTGAAGAGGAGGAGGAGGAGTGA 880
DB 953 GGAAGCCCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1012
OY 881 TGGGCTGACGCTCTCTATGCTGAGATCAAGGCTCCCAAGTTCAATGATCCG 940
DB 1013 TCAAGTGACACCTCTGCTCATGCTGATCATGAGCTCCCAAGTGAATGAGACCCC 1072
OY 941 GCGTACGCGGCTGCTGAGGAGTCAACACAGAGCGGCTGAGCATTTGCGAGGCTGT 1000
DB 1073 GATTGGCAAGGCTGCTGAGAGTGCACACAGAGCGGCTGAGCATGAGGCTGT 1132
OY 1001 GGCAGTATGAAAGCAACAGGCTGAGAGTCCCATGCAAGAAATCATCAATGAGG 1060
DB 1133 GCGTTTACATCAAGCAACACAGCTGAGATGGGACAGCGGAGTCAATCAATGCA 1192
OY 1061 ACAAGTATTTCCAGAGATTTTGTATGTCCTCCGCTGAAAGTTTGTGATTTCCGAC 1120
DB 1193 ACCGTACTTCCGAGATCTTCAATTTGAGGCGAGCTCGGTTTCTCGAGATTTCCATGA 1252
OY 1121 GCGTACAGCGCTGCTATTTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1180
DB 1253 AGCTGCGAGGTTGCTGAGAGATCCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1312
OY 1181 ACCCTTCAAGCAAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1240
DB 1313 ACCCTTCAAGCAAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1372
OY 1241 TAAAGGGGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1300
DB 1373 TGAAGGCGCAATGAGAGATTTTCTGCGCTTACCAACCAATGAGAGAGATGCGCTCC 1432
OY 1301 TGGACAGTAAATGATGATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1360
DB 1433 TTGATGTCAAGATCCATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1492
OY 1361 TCAATGTAAGCTTCTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1420
DB 1493 TCAATGTAAGCTTCTCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1552
OY 1421 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1480
DB 1553 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1612
OY 1481 AGTTTCAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1540
DB 1613 CTCTTCAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1672
OY 1541 AGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1597

```

DB 1673 ACCAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1729

RESULT 10

ABL02582

ID ABL02582 strand; cDNA; 4138 BP.

XX ABL02582;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 2228.

KW Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

XX Drosophila melanogaster.

PN MO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

XX P-PSDB; ABB58479.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

PS Claim 1; SEQ ID NO 2228; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 4138 BP; 1065 A; 1001 C; 1024 G; 1048 T; 0 other;

Query Match 29.2%; Score 504.2; DB 23; Length 4138;

Best Local Similarity 66.6%; Pred. No. 2.3e-105;

Matches 825; Conservative 0; Mismatches 333; Indels 81; Gaps 4;

```

OY 433 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 492
DB 1649 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1708
OY 493 TCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 552
DB 1709 TCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1768
OY 553 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 612
DB 1769 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1828
OY 613 CTCTATCTTCAACACTTTTAACTCTGAGTGTGAGGAGATTCC---GACGAG 669

```



1829	ATCTTCATCTCGAACA	CATTC	TATCC	CAGAC	AGGAG	CCCA	CAATG	ATGATGG	CAGAGGCG	1888
670	AGCATTCGCTCTGGG	AGCTAC	GGGTGG	AGGG	AGAGCT	CTCTGG	ATGATCC	-----	720	
1889	GCGTTCGCTTTGGG	AATTCG	GGTGG	AGGTCC	CTGT	TAGAG	GATGG	CAGAGGCGAT	1948	
721	-----AGCA	ACAGAGCG	GAAGTTCT	CTCTTTCT	TTCTCA	AGAGTT	TGGTCA	TCAGAGCTG	774	
1949	CCCAATACA	AGAGATCA	AAACG	AAATTT	TCGTCG	TTTTTCA	AGTTCG	TGTATTCAGAGCTG	2008	
775	GACAAAGATCT	TATGGCC	TGACAA	CCACCT	CTGTT	GAGTGG	CATCGG	ACACCCACGACC	834	
2009	GATAAGAACT	GTACG	TCCGG	ACAACCA	CTTGT	TCAGTGG	CATCGCA	CTCACACCA	2068	
835	CAGAGACG	ACGCGTTCC	AGGTGA	AAACG	CTGGG	ACCTG	AGTGTG	CGCTGCA	894	
2069	CAGGACG	CGATTCG	AGTTCG	AGTGA	AGCGG	CCGGCG	ATCGCA	TGTGCGCTGC	2128	
895	CTCCTCAT	GCTGGAC	TACCA	CGCTCC	CCAGTTC	AAACTGG	ATCCC	CGCTAG	954	
2129	CTCCTGCT	TGCTGACT	ACCAG	CGCTCG	AGTTCA	AGCTG	GAAC	CGCGACTG	2188	
955	CTGGGCTG	CACACAC	AGAGCG	CTCAG	CCACTT	TCAGG	CCCTGT	GGCAGTAT	1014	
2189	CTGGCGT	CACACAC	AGAGCG	CCG	TGATCAT	TCCG	CCCTGT	TCGAGTAC	2248	
1015	ACCAACAG	CGTGCAG	AGACTCC	CTATG	ACA	AGGATAC	ATCAAT	GGGGA	1074	
2249	ACGCACA	AGGCTAC	AGGATG	CCACG	AGCGG	AGTAC	ATCAAT	TGCAACA	2308	
1075	CAGATTTT	TGATGCT	CTCC	CGCTCA	AGTTTCT	CGAGAT	TCCC	AGCCCTC	1134	
2309	CAGATATT	CAGTTGCC	ACGCG	ATGAG	TTTGC	CGAGAT	ACGCA	ACGCTCA	2368	
1135	CTATTGCC	CCCCGTG	ACCA	ATGTCAT	CA	ACCATG	TCAT	CA	1194	
2369	CTGCAT	CCGCCG	ATCCG	ATGAT	CAATCA	TTTCAT	---TG	AGCGGTG	2425	
1195	AAGNAG	ACGGTGTCT	ATGATCA	TGAC	GTGG	AGTGG	AGGAC	CCATTA	1254	
2426	AAGCAG	ACTCCCTG	CTACG	ACATCG	ATGTG	GAGTGG	ATGAT	ACGCTCA	2485	
1255	AGCAGCTT	CTCTAT	CCACG	GGCCAA	CCAGC	AGGAGAT	CA	GTCTCT	1314	
2486	AA	CAGCTTCT	GATGAG	CACTG	CGG	CCACAG	AGATCC	AGGGCT	2545	
1315	CATGAG	ACGATGTAG	TCCATA	AAACG	CTCA	AGATCC	AGAGG	ACTTCAT	1374	
2546	CACGAG	ACGGTGG	ACACG	ATCA	ACAGAT	GAAG	ACGA	ACGAGG	2605	
1375	TCCAG	AGACCCCA	AAAGCT	ATGTC	CA	AGACCTG	CTCCG	CTCC	1434	
2606	GCCA	AGGATCC	ACAAT	TGTTTAT	TCA	TGCTG	ATCAT	AGTGA	2665	
1435	G-----	-----	-----	-----	-----	-----	-----	-----	1435	
2666	GTTAGT	TTCTTTT	TTTTC	TTTTC	TTTTC	TTTTC	TTTTC	TTTTC	2725	
1436	-TGAT	GACAGT	TAG	CGGGA	ACCCTG	AA	GAGAG	CGCGG	1494	
2726	CTGAT	GACCA	GATGTAG	CTG	CAATCC	CGG	AGGAG	CGCTCG	2785	
1495	CCCTG	GTCC	CAGGAG	CGCTC	AGT	CTGCT	TCTT	CTACT	1554	
2786	CCAT	TGAC	GCAC	AGGCG	CGTTT	CG	CGTCA	TTCTT	2845	
1555	GAGCT	GAGCAG	TGCTG	TTGTG	CG	CAAC	ACCT	TAGAG	1593	
2846	GAAT	TGGAG	CAGG	CGCTG	GGC	ATAC	GAA	CGGCT	2884	

ID	AAAD53120	standard; DNA; 755 BP.
XX	AAAD53120;	
XX	AC	
XX	XX	
DT	28-MAY-2003	(first entry)
XX	XX	
DE	Human BRG1-associated factor (BAF) 60c3 DNA.	
XX	XX	
KW	Human; nuclear receptor cofactor; proliferator-activated receptor; RXR;	
KW	PPAR-gamma; retinoid X receptor; cell proliferative disorder; infection;	
KW	metabolic disorder; cardiovascular disorder; inflammatory disease; acne;	
KW	neurodegenerative disease; Parkinson's disease; psoriasis; cancer; HIV;	
KW	renal disease; atherosclerosis; Alzheimer's disease; diabetes; vaccine;	
KW	osteoporosis; human immunodeficiency virus; BRG1-associated factor; BAF;	
KW	gene therapy; ds.	
OS	XX	
XX	Homo sapiens.	
XX	XX	
PH	Key	Location/Qualifiers
FT	CDS	1..753
FT	FT	/*tag= a
FT	FT	/product= "BAF60c3 protein"
FT	FT	/transl_except= (pos:598..609, aa:Arg-Gln)
FT	FT	/transl_except= (pos:610..618, aa:Met-Pro)
FT	FT	/note= "No stop codon"
FT	FT	/partial
XX	XX	
PN	WO200294877-A2.	
XX	XX	
PD	28-NOV-2002.	
XX	XX	
PF	23-MAY-2002; 2002WO-IB02939.	
XX	XX	
PR	23-MAY-2001; 2001US-292526P.	
XX	XX	
PA	(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.	
XX	XX	
PI	Auwerx J, Gelman L, Debril M, Rocchi S, Picard F;	
XX	XX	
DR	WPI; 2003-120789/11.	
DR	P-PSDB: AAE34721.	



```

Db      78 GGCCTCCGAGATCCGCTGTGAGCCCGGATGCCCCAACAAGGGGCGCCCATGGGCCCCC 137
Qy      278 GGGCTCCCGTACATGGGAGCCCGCCGCTGGCAACCCGGCTGGCCCCCGGG-ATGGA 336
Db      138 GGGCTCCCGTACATGGGAGCCCGCCGCTGGCAACCCGGCTGGCCCCCGGGCATGGA 197
Qy      337 GCGCGCCGCAAGCGAGCGCGCCCGCC-GGCAGAGCCAGGCAAGAGCGAGGCGCA 395
Db      198 GCGCGCCGCAAGCGAGCGCGCCCGCCCGGAGAGCGAGCAAGAGCGAGGCGCA 257
Qy      396 GCGGAGCCCAACCGCCCGCGGAGCGGAGTGCAGAGAGAGAGATGGCTGACAA 455
Db      258 GCGGAGCCCAACCGCCCGCGGAGCGGAGTGCAGAGAGAGAGATGGCTGACAA 317
Qy      456 AATCTCTCCCTCAAGAGATTCGGAGCTGGTCCCGAGTCCAGGCTTACATGACCTCTT 515
Db      318 AATCTCTCCCTCAAGAGATTCGGAGCTGGTCCCGAGTCCAGGCTTACATGACCTCTT 377
Qy      516 GGCATTGAGAGGAATCGGATCAACATCATGCGGAAGGGGGTGGACATCCAGGAGGC 575
Db      378 GGCATTGAGAGGAATCGGATCAACATCATGCGGAAGGGGGTGGACATCCAGGAGGC 437
Qy      576 TCTGAGAGGCGCCATGAGCAAAAGCGGAAGCTGCAGCTCTATATCTCCAGACTTTTAA 635
Db      438 TCTGAGAGGCGCCATGAGCAAAAGCGGAAGCTGCAGCTCTATATCTCCAGACTTTTAA 497
Qy      636 CCTCGGAGAGTGTGATGCTGAGGATTCGACGCGACAGATTGCTCTCGGAGCTACGGGT 695
Db      498 CCTCGGAGAGCTGATGCTGAGGATTCGACGCGACAGATTGCTCTCGGAGCTACGGGT 557
Qy      696 GGAGGGGAAGCTCTCTGATGATCCGAG 722
Db      558 GGAGGGGAAGCTCTCTGATGATCCGAG 584

```

## RESULT 12

ABT07117  
ID ABT07117 standard; cDNA; 450 BP.

```

XX AC ABT07117;
XX DT 07-NOV-2002 (first entry)
XX DE Human ovarian cancer associated coding sequence SEQ ID NO: 279.
XX KW Human; ovarian cancer; cancer; gene; ss.
XX OS Homo sapiens.
XX PN US2002076715-A1.
XX PD 20-JUN-2002.
XX PF 06-JUN-2001; 2001US-0876889.
XX PR 23-SEP-1998; 98US-0159320.
XX PR 08-FEB-1999; 99US-0246429.
XX PR 16-SEP-1999; 99US-0397787.
XX PA (BENS/) BENSON D R.
XX PA (LODE/) LODES M J.
XX PA (MITC/) MITCHAM J L.
XX PA (KING/) KING G E.
XX OS Homo sapiens.
XX PI Benson DR, Lodes MJ, Mitcham JL, King GE;
XX DR WPI, 2002-598720/64.
XX PT Composition for detecting and treating ovarian cancer, comprises a
XX PT specific polypeptide, polynucleotide, T cell population, or antigen
XX PT presenting cell -
XX

```

PS Example 1; Page 105; 188bp; English.

XX CC The present invention relates to a method of detecting the presence of  
 CC ovarian cancer in a patient, involving detecting ovarian cancer  
 CC associated polynucleotides. The method is not only used to detect the  
 CC presence of cancer, preferably ovarian cancer in a patient, but also is  
 CC used to stimulate and/or expand T cells specific for an ovarian tumour  
 CC protein. The sequences can be used in vaccines used to treat cancer. The  
 CC present sequence is an ovarian cancer associated coding sequence.

XX SQ Sequence 450 BP; 71 A; 170 C; 159 G; 48 T; 2 other;

Query Match 23.8%; Score 410.4; DB 24; Length 450;

Best Local Similarity 98.7%; Pred. No. 3.4e-84;

Matches 444; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

```

Qy      40 GAGCGGCGCGCGCGCTCCGGGCGGGTCCCGGGAGACAGATCTCAGAAATGCCCCCTT 99
Db      1 GAGCGGCGCGCGCGCTCCGGGCGGGTCCCGGGAGACAGATCTCAGAAATGCCCCCTT 60
Qy      100 GGTGCTGAGAGCGCGGTGGCTCCCGGCCAGAGCACGAGGGGGGACTGATGACTTCC 159
Db      61 GGTGCTGAGAGCGCGGTGGCTCCCGGCCAGAGCACGAGGGGGGACTGATGACTTCC 120
Qy      160 AGGTGAGAGACCTCTGCTATGACTCCAGTCTTACAGACCCAGCCAGCTGTAGAG 219
Db      121 AGGTGAGAGACCTCTGCTATGACTCCAGTCTTACAGACCCAGCCAGCTGTAGAG 180
Qy      220 CGCCCCGGGAGTCCGTGTGAGCCCGGATGCCCAACA-GGGGCGCCCATGGGCCCCCGG 278
Db      181 CGCCCCGGGAGTCCGTGTGAGCCCGGATGCCCAACAAGGGGGCGCCCATGGGCCCCCGG 240
Qy      279 GGTCTCCCGTACATGGGAGCCCGCCGCTGCAACCCGGCTGGCCCCCGGG-ATGAG 337
Db      241 GGTCTCCCGTACATGGGAGCCCGCCGCTGCAACCCGGCTGGCCCCCGGGAGCTTGGAG 300
Qy      338 CGCGCCCGCAAGCGAGCGCGCCCGCCCGCC-GGCAGAGCGAGGCAAGAGCGAGGCGCGAG 396
Db      301 CGCGCCCGCAAGCGAGCGCGCCCGCCCGCCCGGAGAGCCAGAGCCAGAGGCGCGAG 360
Qy      397 CCGGAGCCCAACCGCCCGCGGAGCGGAGTGCAGTGCAGAGAGAGAGATGGCTGACAAA 456
Db      361 CCGGAGCCCAACCGCCCGCGGAGCGGAGTGCAGAGAGAGAGAGATGGCTGACAAA 420
Qy      457 ATCTCTCCCTCAAGAGATTCGGAGCTGGTC 486
Db      421 ATCTCTCCCTCAAGAGATTCGGAGCTGGTC 450

```

## RESULT 13

ABX72995  
ID ABX72995 standard; cDNA; 450 BP.

```

XX AC ABX72995;
XX DT 14-MAR-2003 (first entry)
XX DE Human ovarian carcinoma antigen partial cDNA sequence #261.
XX KW Human; cancer detection; ovarian carcinoma antigen; ovarian cancer;
XX KW tumour antigen; tumour; OV2; OV3; OV6; OV9; OV10; OV12; OV14; OV17;
XX KW OV18; OV23; OV24; OV27; OV41; OV54; OV57; gene; ss.
XX OS Homo sapiens.
XX PN US6468758-B1.
XX PD 22-OCT-2002.
XX PF 16-SEP-1999; 99US-0397787.
XX PR 23-SEP-1998; 98US-0159320.
XX PR 08-FEB-1999; 99US-0246429.

```

```
XX (CORI-) CORIXA CORP.
XX Benson DR, Lodes MJ, Mitcham JL, King GE;
XX WPI; 2003-147101/14.
XX
XX Determining presence or absence of cancer in patient by contacting
XX patient sample with oligonucleotide that hybridises to polynucleotide
XX encoding ovarian carcinoma antigen, and detecting amount of
XX hybridisation in sample
XX
XX Example 1; Column 201-202; 152pp; English.
XX
XX The present invention relates to compositions and methods for
XX determining presence or absence of cancer in patient. The method
XX comprising contacting a biological sample with an oligonucleotide
XX that hybridises to partial polynucleotide sequence encoding for
XX human ovarian carcinoma antigen, or its complement. The method and
XX compositions are useful for the therapy and diagnosis of ovarian
XX cancer. The method can be used to identify tumour antigens that
XX are secreted from ovarian carcinoma and/or other tumours. Effective
XX cancer detection is achieved using the method of the invention.
XX ABX72720-ABX73050 represent partial cDNA sequences encoding human
XX ovarian carcinoma antigens.
XX
XX Sequence 450 BP; 71 A; 170 C; 159 G; 48 T; 2 other;
XX
XX Query Match 23.8%; Score 410.4; DB 25; Length 450;
XX Best Local Similarity 98.7%; Pred. No. 3.4e-84;
XX Matches 444; Conservative 0; Mismatches 3; Indels 3; Gaps 3;
XX
QY 40 GAGCGGGCGCGCGCTCCGGCGGGGTCCGGGGAGCAGATCCTCAGAAATGGCCCTT 99
Db 1 GAGCGGGCGCGCGCTCCGGCGGGGTCCGGGGAGCAGATCCTCANNATGGCCCTT 60
QY 100 GGTGCTCAGGCGCGGTGGCTCCGGGCCCGCAGGCCAGGAGGGGCACTGGATCTCC 159
Db 61 GGTGCTCAGGCGCGGTGGCTCCGGGCCCGCAGGCCAGGAGGGGCACTGGATCTCC 120
QY 160 AGGTGACAGACCTTCCATCTATGACTCAGGTCTTTAGCACCACCCACCGTGGTACAG 219
Db 121 AGGTGACAGACCTTCCATCTATGACTCAGGTCTTTAGCACCACCCACCGTGGTACAG 180
QY 220 CGCCCGGGGATGCGCTCTGGAGCCCGGATGCCACCA-GGGGCGCCCATGGGCCCCCG 278
Db 181 CGCCCGGGGATGCGCTCTGGAGCCCGGATGCCACCAAGGGGGCCCATGGGCCCCCG 240
QY 279 GGCTCCCGGTATGCGGAGCGCCCGCGTGGACCCCGCTTGGCCCGCGGG-ATGGAG 337
Db 241 GGCTCCCGGTATGCGGAGCGCCCGCGTGGACCCCGCTTGGCCCGCGGGCATGGAG 300
QY 338 CCGGCCCGCAGCAGCAGCGCCCGCGCC-GGCAGCCAGGCACAGAGCCAGGGCCAG 396
Db 301 CCGGCCCGCAGCAGCAGCGCCCGCGCCCGCGGCGCAGAGCCAGGCACAGAGCCAGGGCCAG 360
QY 397 CCGGAGCCACCGCCCGCGCGGAGCGCAGTGCACAGAGAGGAAGATGGCTGACAAA 456
Db 361 CCGGTGCCACCGCCCGCGCGGAGCCGCGAGTGCACAGAGAGGAAGATGGCTGACAAA 420
QY 457 ATCTCTCCCTCAAGGATTCGGGAGCTGGTC 486
Db 421 ATCTCTCCCTCAAGGATTCGGGAGCTGGTC 450
XX
RESULT 14
ABX47225
ID ABX47225 standard; cDNA; 285 BP.
XX
AC ABX47225;
XX
XX 21-FEB-2003 (first entry)
XX
```

```
DE Bovine EST associated with lactation/muscle/fat deposition #12390.
XX Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
XX muscle deposition; fat deposition; genome mapping; gene identification;
XX gene analysis; cattle breeding.
XX Bos Taurus.
XX OS
XX US2002137139-A1.
XX PN
XX 26-SEP-2002.
XX PD
XX 24-SEP-2001; 2001US-0960352.
XX PF
XX 12-JAN-1999; 99US-115707P.
XX PR
XX 11-JAN-2000; 2000US-0480902.
XX PA (BYAT/) BYATT J C.
XX PA (MATH/) MATHIALAGAN N..
XX PA (TAON/) TAO N.
XX PA (WARR/) WARREN W C.
XX
XX Byatt JC, Mathialagan N, Tao N, Warren WC;
XX WPI; 2003-110599/10.
XX
XX New nucleic acid associated with lactation, and muscle and fat
XX deposition, useful for genome mapping, gene identification and
XX analysis, cattle breeding, or for genetically improving cattle
XX Claim 2; SEQ ID No 12390; 245pp; English.
XX
XX The invention relates to a purified nucleic acid molecule associated with
XX lactation or muscle and fat deposition (designated LMFD), derived
XX from cattle, and the LMFD nucleic acid can specifically hybridise to a
XX second nucleic acid molecule comprising any of 1512 nucleotide
XX sequences, appearing as ABX34836-ABX49947, or complements of them.
XX Also included are; (1) a transformed cell having a nucleic acid
XX comprising an LMFD nucleic acid linked to a promoter and a 3' non-
XX translated sequence that functions in the cell to cause termination of
XX transcription and addition of polyadenylated ribonucleotides to a 3' end
XX of the mRNA molecule; and (2) determining a level or pattern of a
XX molecule in a bovine cell or tissue comprising: (a) incubating a marker
XX nucleic acid (comprising any of the 1512 nucleic acid sequences or its
XX complement or fragment) with a complementary nucleic acid molecule
XX obtained from the bovine cell or tissue, where hybridisation between the
XX marker nucleic acid and the complementary nucleic acid permits the
XX detection of the molecule; and (b) detecting the level or pattern of the
XX complementary nucleic acid, where the detection of the complementary
XX nucleic acid is predictive of the level or pattern of the molecule.
XX The LMFD nucleic acid is used for determining a level or pattern
XX of a molecule in a bovine cell or tissue. It is useful for genome
XX mapping, gene identification and analysis, cattle breeding, preparation
XX of constructs for use in cattle gene expression, or for genetically
XX improving cattle. The present sequence is one of the 1512 bovine
XX LMFD EST (expressed sequence tag) nucleic acids.
XX Note: The present sequence was not shown in the specification but
XX was obtained in electronic format from the USPTO web site:
XX seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
XX Sequence 285 BP; 62 A; 79 C; 76 G; 68 T; 0 other;
XX
XX Query Match 12.0%; Score 206.2; DB 25; Length 285;
XX Best Local Similarity 83.0%; Pred. No. 1.4e-37;
XX Matches 235; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
XX
QY 559 GTGGACATCCAGGAGGCTCTGAAGAGGCCATGAAGCAAAAGCGAAAGCTCGACTCTAT 618
Db 3 GTGGTCATTGCTGATGCTCTGAAGCTGTTTACGAAGCAGAGCGCAAGCTGGTGGCTAT 62
QY 619 ATCTCCACACTTTTAACCCCTGCGAAGTCTGATCTGAGATTCGACGGCAGCATTCGCC 678
Db 63 ATCTCCAATACCTTTTAACCCCTGCTGAAGCCCGATCGCTAGGATTCGACGGCAGCATTCGCC 122
```

QY 679 TCCTGGAGCTACGGGTGAGGGGAAAGCTCTGGATGATCCAGCAAGAGCGGAG 738  
 DB 123 TCCTGGAGAGCTGCTGTGAGGGGAAAGCTCTGGATGATCCAGCTATGCGGAG 182  
 QY 739 TTCTCTCTCTCTCTCAAGATTGTGATCGAGCTGGACAAAGATCTTTATGSCCTGAC 798  
 DB 183 TTCTGCTCTCTCTCAAGAGCTGCTGATGATCTGGACATGACCTTTATGSCCTGAC 242  
 QY 799 AACCACTCTGTTGAGTGGCATCGGACACCCAGCAGCAGGAGA 841  
 DB 243 AACCACTCTGTTGAGTGGCATCGGACACCCAGCAGCAGGAGA 285

RESULT 15  
 AAZ80712/c  
 ID AAZ80712 standard; cDNA; 452 BP.  
 AC AAZ80712;  
 XX  
 DT 07-APR-2000 (first entry)  
 XX  
 DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:796.  
 XX  
 KW Human; gene expression product; diagnosis; tumour; colon cancer;  
 KW colorectal adenocarcinoma; cell line SW480; cell proliferation;  
 KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;  
 KW hyperplasia; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9964576-A2.  
 XX  
 PD 16-DEC-1999.  
 XX  
 PF 09-JUN-1999; 99WO-IB01062.  
 XX  
 PR 10-JUN-1998; 98US-0088801.  
 XX  
 PA (FARB ) BAYER CORP.  
 XX  
 PI Endege WO, Steinmann KE, Aslie JH, Burgess CC, Bushnell SE;  
 PI Carroll E, Carino TJ, Dertl A, Ford DM, Lewis ME, Monahan JE;  
 PI Schlegel R;  
 XX  
 DR WPI; 2000-087220/07.  
 XX  
 PT Novel nucleic acids, used to develop products for the diagnosis and  
 PT treatment of disorders involving unwanted cell proliferation,  
 PT particularly cancers, especially colon cancer -  
 XX  
 PS Claim 15; Page 448; 469pp; English.  
 XX  
 CC AAZ79917 to AAZ80766 represent double stranded cDNA clones isolated from  
 CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The  
 CC cDNA clones can be used to generate antisense oligonucleotides which  
 CC can be used for antisense therapy. Methods and products from the present  
 CC invention can be used for identifying and/or classifying cancerous cells  
 CC present in a human tumour, particularly in solid tumours, e.g.  
 CC carcinomas and sarcomas, e.g. breast or colon cancers. The cDNA clones  
 CC can be used for developing agents for the diagnosis and treatment of  
 CC disorders involving unwanted cell proliferation, such as neoplasia,  
 CC dysplasia or hyperplasia.  
 CC  
 SQ Sequence 452 BP; 99 A; 124 C; 104 G; 122 T; 3 other;

Query Match 11.9%; Score 205; DB 21; Length 452;  
 Best Local Similarity 73.4%; Pred. No. 3.1e-37;  
 Matches 312; Conservative 0; Mismatches 90; Indels 23; Gaps 3;

QY 480 GCTGCTCCCGAGTCCAGGCTTACATGAGCCTTGGCATTTGAGGAAACTGGATCA 539  
 DB 424 GCAGTACCAAGATCCAGGCTATGATCTTGGCTTTGAAAGG-ACTGAGCA 366

QY 540 AACCATCATGCGAAGGGGGTGGACATCCAGAGGCTGTAAGAGGCCCATGAGCAAAA 599  
 DB 365 GACTATCATAGAGAAC-GGCTTGATATCCAAAGGCTTTGAAAGCTCCATCAAGCAAAA 307  
 QY 600 GCGGAAGTGCAGACTCTATATCTCCAGACCTTTAAACCTGCGAAGTGTGATGAGGA 659  
 DB 306 ACGGAAGTGGAAATTTTCATTTCTTAACCTTTCAATCCGCTTAAGTGAATGCCGAGA 247  
 QY 660 TTCCGACGCGACATTTGCTCTGGAGTACGGGTGAGGGGAAGCTCTGG----- 712  
 DB 246 TGGGGAAGGAGCGGTGGCTTCTGGGAGCTTGGGTAGAGGACGCTCTGAGAGATTTC 187  
 QY 713 -----ATGATCCAGCAAAAGAGAGGGAAGTTCTGCTTTCTTCAAGAG 758  
 DB 186 AGCTTTGTCAAATATATGATGCCATTAACAAAGAGAAATCTCTCTTTTAAATTC 127  
 QY 759 TTTGATCATGAGCTGAGCAAAAGATCTTTATGSCCTGACAAACCACTGTGATGAGGA 818  
 DB 126 CTTGATGATGAACTGAGCAAAAGACTGTATGGGCAAGACAAACATCTGTGATGAGGA 67  
 QY 819 TCGAACCCACGACCCAGAGAGCGAGCGCTTCCAGGTGAAAAGGCTGSGGACCTGAG 878  
 DB 66 CAGGACCGGCACTACCCAGAGAGCCGATGCTTTCAGGTGAAAGCGGCGGAGAGCTGAA 7  
 QY 879 TGTGC 883  
 DB 6 TGTAC 2

Search completed: October 11, 2003, 09:23:56  
 Job time : 367 secs

**THIS PAGE BLANK (USPTO)**

```

RESULT 1
US-09-397-787-279
Sequence 279, Application US/09397787
Patent No. 6468758
GENERAL INFORMATION:
APPLICANT: Benson, Darin R.
APPLICANT: Lodes, Michael J.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
FILE REFERENCE: 210121.466C2
CURRENT APPLICATION NUMBER: US/09/397,787
CURRENT FILING DATE: 1999-09-16
NUMBER OF SEQ ID NOS: 334
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 279
LENGTH: 450
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(450)
OTHER INFORMATION: n = A,T,C or G
US-09-397-787-279

Query Match          23.8%; Score 410.4; DB 4; Length 450;
Best Local Similarity 98.7%; Pred. No. 2,2e-90;
Matches 444; Conservative 0; Mismatches 3; Indels 3; Gaps 3

QY      40  GAGCGGGCGCGCGCTCCGGGCGGGGTCCGGGGAGACAGATCTTCAGAAATGGCCCTT 99
Db      1  GAGCGGGCGCGCGAGCTCCGGGCGGGGTCCGGGGAGACAGATCTTCAGAAATGGCCCTT 60

QY      100 GGTGTGCAAGGCGCGGTGGGCTCCGGGGCCAGGACCGAGGGGGCAATTGATCTTCC 159
Db      61  GGTGTGCAAGGCGCGGTGGGCTCCGGGGCCAGGACCGAGGGGGCAATTGATCTTCC 120

QY      160 AGGTGAGAGACCTTCACATATGACTCCAGGCTTTGAGCACCACCAACCGTGTAAG 219
Db      121  AGGTGAGAGACCTTCACATATGACTCCAGGCTTTGAGCACCACCAACCGTGTAAG 180

QY      220 CGGCGCGGAGTCCGCTTGAGAGCCCGGATGCCACCA-GGGGGCCCATAGGGCCCCCG 278
Db      181  CGGCGCGGAGTCCGCTTGAGAGCCCGGATGCCACCAAGGGGGGCCCATAGGGCCCCCG 240

QY      279 GGTCTCCCGTACATAGGGAGACCCCGCGGTGACCCGAGCCCGGCCCCCGGAG-ATGAG 337
Db      241  GGTCTCCCGTACATAGGGAGACCCCGCGGTGACCCGAGCCCGGCCCCCGGAGATGAG 300

```









Db	Qy	Db	Qy
673 675	CGCGCGTCCCCCACTTTTTCGCGCGGTAACGCCCAACTTCGCGCGGTGCGCCGCGGAGTGCCCG	673 734	
	316		375
	GCCTGACCCCCCGGGGATGAGACCCGCGCGCAAGAGACGCGCCCCCGCGGCAGAGC		
673 735	TCCGCGCCCCCGCTGGAAACCGTTGACACCGTCGCTGCGGACCTTCAGTCCCGCCGACG	673 794	
	376		420
	CAGGCACAGAGCCAGGCGCCAGCCCGAGACCCAGCCCGCCCGCGGG		
673 795	CCGCGGTGACCGCCGCGCCCGCGGCGGCACCGCTTGCCCGCCGACG	673 839	

```

1 RESULT 10
2 US-07-928-611-12
3 Sequence 12, Application US/07928611
4 Patent No. 5569601
5 GENERAL INFORMATION:
6 APPLICANT: Van Tol, Hubert H.M.
7 APPLICANT: Civeilli, Olivier
8 TITLE OF INVENTION: A No. 5569601el Human Dopamine Receptor and Uses
9 NUMBER OF SEQUENCES: 22
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Allegretti & Witcoff, Ltd.
12 STREET: 10 South Wacker Drive, Suite 3000
13 CITY: Chicago
14 STATE: Illinois
15 COUNTRY: USA
16 ZIP: 60606
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: Patentin Release #1.0, Version #1.25
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/07/928,611
24 FILING DATE: 19920810
25 CLASSIFICATION: 435
26 ATTORNEY/AGENT INFORMATION:
27 NAME: No. 5569601nan, Kevin E
28 REGISTRATION NUMBER: 35,303
29 REFERENCE/DOCKET NUMBER: 90,1092-B
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 312-715-1000
32 TELEFAX: 312-715-1234
33 TELEX: 810-221-8317
34 INFORMATION FOR SEQ ID NO: 12:
35 SEQUENCE CHARACTERISTICS:
36 LENGTH: 803 base pairs
37 TYPE: NUCLEIC ACID
38 STRANDEDNESS: single
39 TOPOLOGY: linear
40 MOLECULE TYPE: DNA (genomic)
41 FEATURE:
42 NAME/KEY: exon
43 LOCATION: 1..803
44 IDENTIFICATION METHOD: experimental
45 OTHER INFORMATION: /evidence=EXPERIMENTAL
46 OTHER INFORMATION: /standard name="Alternate Exon 3: D4.7"
47 OTHER INFORMATION: /note="This sequence represents the third exon of
48 OTHER INFORMATION: allele D4.7 of the human D4 dopamine receptor
49 OTHER INFORMATION: gene"
50 FEATURE:
51 NAME/KEY: misc feature
52 LOCATION: 257..262
53 IDENTIFICATION METHOD: experimental
54 OTHER INFORMATION: /function="PstI site"
55 OTHER INFORMATION: /evidence=EXPERIMENTAL
56 OTHER INFORMATION: /standard name="PstI site"
57 OTHER INFORMATION: /label="PstI
58 OTHER INFORMATION: /note="This sequence is a PstI site whereby
59 OTHER INFORMATION: digestion of human genomic DNA produces a RFLP"
60 FEATURE:
61 NAME/KEY: repeat region
62 LOCATION: 346..682

```

```

? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /rpt type= "tandem"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /rpt unit= 346 .. 394
? OTHER INFORMATION: /note= "This sequence is a repeat found in 7 known
? OTHER INFORMATION: alleles of the human D4 dopamine receptor gene
? OTHER INFORMATION: encoding a 16 amino acid sequence repeated 7 times
?
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 2..803
?
US-07-928-611-12

```

```

Query Match      2.6%; Score 44.8; DB 1; Length 803;
Best Local Similarity 47.2%; Pred. No.0.11;
Matches 202; Conservative 0; Mismatches 222; Indels 4; Gaps 2

QY      CGAGCGCAGAGCGCCCGAGCCGAGCCCGCAGCAGGAGCGGAGCGGCGCGCTCCGAGCGCG 66
Db      281 CGGCGCAAGCTGCACGCGCGCGCGCCCGCGCCGACCGCAGCGGCGCTCGCGCTTCCGCC 340
QY      67 GTCCCGGAGGAGCAGATCTCAGATGGCCCTTGCTGTCGAGGCGCGGTGGCTTCGCG 126
Db      341 ACGCCACCCGCGCGCCCGCTTCCCGCAGGACCCCTGCGGCGCCGACTGTGCGCCCCCGCG 400
QY      127 CCGAGGCGACGAGGGGGGACATCGATGACTCTCCAGTGTGAGAGACCCCTGCATCTATGACT 186
Db      401 CCGGCGCTTCCCGCGGGGCTCTGTGCGGCCCGAC-TGTGCGCCCGCGCGCGCGCGCGCTCC 459
QY      187 CCAGGTTCTTGACAGCACCACCGCTGGTACAGCGCCCGGAGTGCCTTGGAGCCGCG 246
Db      460 CCGCGAGACCCCTGCGCGGCCCGCAGCTGTGCGCCCCCGCGCGCGCTCCCGCAGAGACCTCTG 519
QY      247 ATGCGCCACGAGGGGCGCCATGGGCCCCCGCGGCTCCCGCTACATGGGACGCCGCG 306
Db      520 CGGCGCCGACTGTGCGCCCCCGCGCGCGCGCGCTTCCCGGGGTCCTCGCGCGCCGACTG 579
QY      307 TGGCAGCC--GGCTTGCGCCCCCGCGGGATGGAGCCCGCCCGCAGAGCAGCAGCGCCCC 363
Db      580 TGCACCCCCCGCGCGCCCGCTTCCCGCAGGACCCCTGCGGCCCCGACTGTGCGCCCCCGCG 639
QY      364 GCCCGGAGAGCGCAGGCGCAGAGGCGAGGCGCAGCGCCGACCGCGCCCGCGCGAGC 423
Db      640 GCGCGGCTTCCCCCGGAGCCCTCGGCTTCCCACTGTGCTTCCCCCGACGCGCGTAGAGC 699
QY      424 CGCAGTGC 431
Db      700 CGCGCGCG 707

RESULT 11
US-08-487-811A-12
; Sequence 12, Application US/08487811A
; Patent No. 588326
; GENERAL INFORMATION:
; APPLICANT: Civeili, Olivier
; APPLICANT: Van Tol, Hubert H.M.
; TITLE OF INVENTION: A No. 5883226e1 Human Dopamine Receptor and Uses
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,811A
; FILING DATE: 07-JUN-1995
;

```



```

FEATURE:
NAME/KEY: repeat region
LOCATION: 346..682
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /rpt_type="tandem"
OTHER INFORMATION: /evidence=EXPERIMENTAL
OTHER INFORMATION: /rpt_unit=346..394
OTHER INFORMATION: /note="This sequence is a repeat found in 7 known
OTHER INFORMATION: alleles of the human D4 dopamine receptor gene
OTHER INFORMATION: encoding a 16 amino acid sequence repeated 7 times
FEATURE:
NAME/KEY: CDS
LOCATION: 2..803
US-09-060-694-12

Query Match
Best Local Similarity 47.2%; Score 44.8; DB 3; Length 803;
Matches 202; Conservative 0; Mismatches 222; Indels 4; Gaps 2;

QY 7 CGGCGCAGGCGCGCGAGCGCGGAGCGAGGAGCGGCGCGCGCTCCGGGCGGG 66
DB 281 CGCGCCAGCTGCACGCGCGCGCGCGCGCGCGCGAGCCAGCGCGCGCTTCCCGC 340
QY 67 GTCCCGGGGAGCAGATCCTCAGAAATGCGCTTGCTGTCAGAGCGCGGTGGGCTCCGGG 126
DB 341 ACGCCACCGCGCGCGCGCGCTCCCGCGAGACCCCTGGGCGCGCGAGCTGCGCGCGCGG 400
QY 127 CCCAGGACCGAGGGGGGCACTGAGTACTCTCCAGGTGACAGACCCCTGCATATGACT 186
DB 401 CCCGCGCTTCCCGGGGTCTGCGCGCGCGCGAG-TCGTGCGCGCGCGCGCGCGCTCC 459
QY 187 CGAGCTTTCAGACCCAGCCAGCCGTGTAGACCGCGCGCGGATGCGCTTGAAGCGCG 246
DB 460 CCGGACCCCTGCGCGCGCGAGCTGCGCGCGCGCGCGCGCGCTCCCGAGACCCCTG 519
QY 247 ATGCCCGACGAGGGGGGCGCATGAGGCGCGCGCGCTCCCGTACATGAGGACCGCGCG 306
DB 520 CGGCGCGAGCTGTGCGCGCGCGCGCGCGCGCGCTTCCCGGGGTCTGCGCGCGCGAG 579
QY 307 TCGACCC--GCGCTGCGCGCGCGCGAGTGAAGCGCGCGCGCGAGCGAGCGCGCG 363
DB 580 TCGCGCGCGCGCGCGCGCGCTCCCGAGACCCCTGCGCGCGCGAGCTGCGCGCGCG 639
QY 364 GCGCGCGAGCGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 423
DB 640 GCGCGCGCTCCCGCGAGCCCTGCGGTCTCAACTGTGCTCCCGCGAGCGCGTGAAGC 699
QY 424 CGCAGTGC 431
DB 700 CGCGCGCG 707

RESULT 13
US-09-378-074-12
Sequence 12, Application US/09378074
Patent No. 6437114
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
Civelli, Olivier
TITLE OF INVENTION: A No. 6437114el Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/378,074
FILING DATE: 20-Aug-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/928,611
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: No. 6437114nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 90,1092-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 1..803
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence=EXPERIMENTAL
/standard_name="Alternate Exon 3: D4.7"
/note="This sequence represents the third exon of
allele D4.7 of the human D4 dopamine receptor
gene"
FEATURE:
NAME/KEY: misc-feature
LOCATION: 257..262
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function="Psi site"
/evidence=EXPERIMENTAL
/standard_name="Psi site"
/label=Psi
/note="This sequence is a Psi site whereby
digestion of human genomic DNA produces a RFLP"
FEATURE:
NAME/KEY: repeat region
LOCATION: 346..682
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /rpt_type="tandem"
/evidence=EXPERIMENTAL
/rpt_unit=346..394
/note="This sequence is a repeat found in 7 known
alleles of the human D4 dopamine receptor gene
encoding a 16 amino acid sequence repeated 7 times
FEATURE:
NAME/KEY: CDS
LOCATION: 2..803
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-378-074-12

Query Match
Best Local Similarity 47.2%; Score 44.8; DB 4; Length 803;
Matches 202; Conservative 0; Mismatches 222; Indels 4; Gaps 2;

QY 7 CGGCGCAGGCGCGCGAGCGCGGAGCGAGGAGCGGCGCGCGCTCCGGGCGGG 66
DB 281 CGCGCCAGCTGCACGCGCGCGCGCGCGCGCGCGAGCCAGCGCGCGCTTCCCGC 340
QY 67 GTCCCGGGGAGCAGATCCTCAGAAATGCGCTTGCTGTCAGAGCGCGGTGGGCTCCGGG 126
DB 341 ACGCCACCGCGCGCGCGCGCTCCCGCGAGACCCCTGGGCGCGCGAGCTGCGCGCGG 400
QY 127 CCCAGGACCGAGGGGGGCACTGAGTACTCTCCAGGTGACAGACCCCTGCATATGACT 186
DB 401 CCCGCGCTTCCCGGGGTCTGCGCGCGCGCGAG-TCGTGCGCGCGCGCGCGCGCTCC 459

```

```
QY 187 CCAGGTCTTTCAGACCCACCCACCGTGGTACAGCGCCCCCGGGATGCCCTCTGGAGCCCGG 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 CCCGACCCCTGGGCCCCGAGTGTGGCCCCCGCGCCCGGCTCCCGCCAGGACCCCTG 519
QY 247 ATGCCCCACAGGGGGCGCCATGGGCCCCCGGGCTCCCGTACATGGGAGAGCCCGCGG 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 CGGCCCCAGACTGTGGCCCCCGCGCCCGGCTTCCCGGGGTCCCTCGGCCCGGACTG 579
QY 307 TGGGACCC---GGCCTGGCCCCCGGGATGGAGCCCGCCGCAAGCCAGAGCGCCCC 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 TGGCGCCCCCGCGCCCGGCTCCCCCAGAGACCCCTGGCGGCCCGGACTGTGGCCCCCGG 639
QY 364 GCCCGCAGAGCCAGGCGCAGAGCCAGGGCCAGCGCCAGCCCGCCCCCGCGGAGC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 GCCCGGCTCCCCCGGACCCCTGGGCTCCAACCTGTCTCCCCCGGACGCGTCAAGAGC 699
QY 424 CGCAGTGC 431
    ||| ||| |||
Db 700 CGCCGCGC 707

RESULT 14
PCT-US93-07370-12
; Sequence 12, Application PC/TUS9307370
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: A Novel Human Dopamine Receptor and Uses
; NUMBER OF SEQUENCES: 22
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07370
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1..803
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "Alternate Exon 3: D4.7"
; OTHER INFORMATION: /note= "This sequence represents the third exon of
; OTHER INFORMATION: allele D4.7 of the human D4 dopamine receptor
; OTHER INFORMATION: gene"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 257..262
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "psti site"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "Psti site"
; OTHER INFORMATION: /label= Psti
; OTHER INFORMATION: /note= "This sequence is a Psti site whereby
; OTHER INFORMATION: digestion of human genomic DNA produces a RFLP"
; FEATURE:
; NAME/KEY: repeat_region
; LOCATION: 346..682
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /rpt_type= "tandem"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /rpt_unit= 346..394
; OTHER INFORMATION: /note= "This sequence is a repeat found in 7 known
; OTHER INFORMATION: alleles of the human D4 dopamine receptor gene
; OTHER INFORMATION: encoding a 16 amino acid sequence repeated 7 times
; NAME/KEY: CDS
```

```
; LOCATION: 2..803
PCT-US93-07370-12
Query Match 2.6%; Score 44.8; DB 5; Length 803;
Best Local Similarity 47.2%; Pred. No. 0.11; 22; Indels 4; Gaps 2;
Matches 202; Conservative 0; Mismatches 0;

QY 7 CGGCGCAGGGCGCCGAGCCGAGCCGAGCAGGAGGGCGCGGCTCCGGGCGCCGG 66
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 CGGCCCAAGCTGCACGGCGCGCGCCCGCGACCCAGCGGCCCTTGGCCCGCTTCCCC 340
QY 67 GTCCCGGGGAGCAGATCCTCAGAAATGGCCCTTGGTGTGCAGCGCGGGTGGGCTCCGG 126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 341 ACGCCACCCGCGCGCCCGCTCCCGCCAGGACCCCTTGGCGGCCCGGACTGTGCGCCCG 400
QY 127 CCAGGACACGAGGGGGCACTGGATGACTCTCCAGGTGCAGGACCCCTGCCATCTATGACT 186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 401 CCCGGCCTTCCCCGGGGTCTCTGGCGGCCCGAC--TGTGGCGCCCGCGGCCCTCCC 459
QY 187 CCAGGTCTTTCAGCACCCACCCACCGTGGTACAGCGCCCCCGGGATGCCCTCTGGAGCCCGG 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 CCCGACCCCTGGGCCCCGAGTGTGGCCCCCGCGCCCGGCTCCCGCCAGGACCCCTG 519
QY 247 ATGCCCCACAGGGGGCGCCATGGGCCCCCGGGCTCCCGTACATGGGAGAGCCCGCGG 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 520 CGGCCCGGACTGTGGCCCCCGCGCCCGGCTTCCCGGGGTCCCTCGGCCCGGACTG 579
QY 307 TGGGACCC---GGCCTGGCCCCCGGGATGGAGCCCGCCGCAAGCCAGAGCGCCCC 363
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 580 TGGCGCCCCCGCGCCCGGCTCCCCCAGAGACCCCTGGCGGCCCGGACTGTGGCCCCCGG 639
QY 364 GCCCGCAGAGCCAGGCGCAGAGCCAGGGCCAGCGCCAGCCCGCCCCCGCGGAGC 423
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 GCCCGGCTCCCCCGGACCCCTGGGCTCCAACCTGTCTCCCCCGGACGCGTCAAGAGC 699
QY 424 CGCAGTGC 431
    ||| ||| |||
Db 700 CGCCGCGC 707

RESULT 15
US-09-016-434-1276
; Sequence 1276, Application US/09016434
; Patent No. 650938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
```

```

: REFERENCE/DOCKET NUMBER: PA-0002 US
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (650) 855-0555
:
: TELEFAX: (650) 845-4166
:
: INFORMATION FOR SEO ID NO: 1276:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 1504 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
:
: TOPOLOGY: linear
:
: IMMEDIATE SOURCE:
:
: LIBRARY: GENBANK
:
: CLONE: G291945
:
US-09-016-434-1276

```

[illegible]

Search completed: October 11, 2003, 11:20:04  
Job time : 117 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 10:35:17 : Search time 351 seconds  
(without alignments)  
12744.885 Million cell updates/sec

Title: U66619  
Perfect score: 1724  
Sequence: 1 GAATTCGGCGGCGAGCGGCC.....GTCAATCATCTGGGAATTC 1724

Scoring table:  
Gapop 10.0 , Gapext 1.0

Searched: 1731049 seqs, 1297405648 residues

Total number of hits satisfying chosen parameters: 3462098

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	450.8	26.1	485	11	US-09-918-995-16164
2	421.2	24.4	462	11	US-09-918-995-28488
3	412.2	23.9	474	11	US-09-918-995-26465
4	410.4	23.8	450	9	US-09-876-889-279
5	346	20.1	447	11	US-09-918-995-28419
6	340.4	19.7	471	11	US-09-918-995-28605
7	206.2	12.0	285	10	US-09-960-352-12390
8	203	11.9	452	10	US-09-879-536-796
9	203	11.8	495	11	US-09-918-995-2832
10	180.4	10.5	420	11	US-09-918-995-34033
11	179.2	10.4	690	10	US-09-879-536-645
12	170.4	9.9	8788	10	US-09-764-877-2295
13	170.4	9.9	8788	10	US-09-860-670-247
14	162	9.4	440	11	US-09-918-995-27234
15	137.8	8.0	310	14	US-10-070-676-23
16	132	7.7	520	9	US-09-864-761-8048

C	17	129.2	7.5	219	9	US-09-864-761-24793	Sequence 24793, A
	18	91.6	5.3	15500	9	US-09-764-869-2109	Sequence 2109, Ap
	19	91.6	5.3	15500	14	US-10-091-504-2109	Sequence 2109, Ap
	20	84.6	4.9	2017	14	US-10-102-806-137	Sequence 137, App
	21	82.8	4.8	508	11	US-09-918-995-997	Sequence 997, App
	22	67.8	3.9	215	10	US-09-867-701-9486	Sequence 9486, Ap
	23	66.2	3.8	740	9	US-09-770-149-60	Sequence 60, Appl
	24	60	3.5	60	12	US-09-908-975-6179	Sequence 6179, Ap
	25	53	3.1	9025608	14	US-10-156-761-1	Sequence 1, Appl
	26	50.6	2.9	1437	14	US-10-128-714-2188	Sequence 2188, Ap
	27	50.6	2.9	1437	14	US-10-128-714-2188	Sequence 7188, Ap
	28	50.6	2.9	1493	14	US-10-128-714-1188	Sequence 1188, Ap
	29	50.6	2.9	1493	14	US-10-128-714-1188	Sequence 6188, Ap
	30	50.6	2.9	3340	14	US-10-128-714-187	Sequence 187, App
	31	50.6	2.9	3340	14	US-10-128-714-187	Sequence 5187, App
	32	50.6	2.9	3493	14	US-10-128-714-188	Sequence 188, App
	33	50.6	2.9	3493	14	US-10-128-714-188	Sequence 5188, App
	34	50.2	2.9	4908	13	US-10-001-887-33	Sequence 33, Appl
	35	49.8	2.9	28000	12	US-10-091-625-11	Sequence 11, Appl
	36	49.8	2.9	28000	12	US-10-096-399A-11	Sequence 11, Appl
	37	49.4	2.9	522	13	US-10-027-632-129532	Sequence 129532, App
	38	48.6	2.8	2112	14	US-10-219-449-3	Sequence 3, Appl
	39	48.6	2.8	2154	14	US-10-219-449-1	Sequence 1, Appl
	40	47.2	2.7	2678	12	US-10-017-161-1647	Sequence 1647, Ap
	41	47	2.7	2561	10	US-09-976-740-48	Sequence 48, Appl
	42	47	2.7	2561	13	US-10-023-529-48	Sequence 48, Appl
	43	47	2.7	2561	13	US-10-023-523-48	Sequence 48, Appl
	44	46.6	2.7	154746	12	US-09-827-688-8	Sequence 8, Appl
	45	46.6	2.7	154746	12	US-09-827-688-8	Sequence 8, Appl

## ALIGNMENTS

```
RESULT 1
US-09-918-995-16164
; Sequence 16164, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16164
; LENGTH: 485
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(485)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-16164

Query Match
Best Local Similarity 98.5% Pred. No. 5.6e-120;
Matches 455; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 710 TGGATGATCCAGCAAGAGGAGTTCTTCTTTCTTCAAGATTGTCATCG 769
    |||
DB 23 TGGAGAACTGACGACAGAGGAGTTCTTCTTTCTTCAAGATTGTCATCG 82
    |||
QY 770 AGCTGACAAAGATTATATGCGCTTGCAACACCTGTTGAGTGCATCGGAACCA 829
    |||
DB 83 AGCTGACAAAGATTATATGCGCTTGCAACACCTGTTGAGTGCATCGGAACCA 142
    |||
QY 830 CGACCAGAGAGCGAGCGCTTCCAGTGAACGCGCTGGGACCTGAGTGGCCTCA 889
    |||
```

Db 143 CGACCCAGGAGCGAGCGCTTCCAGGTGAACCGCGCTGGGACCTGAGTGTGCCTGCA 202  
Qy 890 CCTGCTCTCTATGCTGGAATACAGCGCTCCCGAGTTCAAACTGGATCCCGCTTAGGCC 949  
Db 203 CGCTGCTCTCTATGCTGGAATACAGCGCTCCCGAGTTCAAACTGGATCCCGCTTAGGCC 262  
Qy 950 GGCTGCTGGGCTGCACACACAGAGCGCTGACGATTTGTCCAGCGCTGTGGGAGTATG 1009  
Db 263 GGCTGCTGGGCTGCACACACAGAGCGCTGACGATTTGTCCAGCGCTGTGGGAGTATG 322  
Qy 1010 TGAACACCAACAGGCTGCAGGACTCCCATGACAAGGAATACATCAATGGGACAGTATT 1069  
Db 323 TGAACACCAACAGGCTGCAGGACTCCCATGACAAGGAATACATCAATGGGACAGTATT 382  
Qy 1070 TCACGACAGATTTTGAATGTCCCGCTGAAAGTTTCTGAGATTTCCCGAGCGCTCACAG 1129  
Db 383 TCACGACAGATTTTGAATGTCCCGCTGAAAGTTTCTGAGATTTCCCGAGCGCTCACAG 442  
Qy 1130 CCTGCTATTGCCCTGACCCCAATGTGTCATCAACCATGTCA 1171  
Db 443 CCTGCTATTGCCCTGACCCCAATGTGTCATCAACCATGTCA 484  
RESULT 2  
US-09-918-995-28488  
; Sequence 28488, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 28488  
; LENGTH: 462  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(462)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-28488

Query Match 24.4%; Score 421.2; DB 11; Length 462;  
Best Local Similarity 99.3%; Pred. No. 2e-111;  
Matches 423; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1183 CCTTCAGACCAAGAGACAGCGCTGCTATGACATTTGAGTGGAGGTGGAGGAGCATTA 1242  
Db 37 CCTTCAGACCAAGAGACAGCGCTGCTATGACATTTGAGTGGAGGTGGAGGAGCATTA 96  
Qy 1243 AAGGGGAGATGAGAGCTTCTCTATCCAGCGCAACACAGAGAGATCAGTCTCTG 1302  
Db 97 AAGGGGAGATGAGAGCTTCTCTATCCAGCGCAACACAGAGAGATCAGTCTCTG 156  
Qy 1303 GACAGTAAGATCCATGAGACGATTTGAGTCCATAAACCAGCTCAAGATCCAGAGGAGCTTC 1362  
Db 157 GACAGTAAGATCCATGAGACGATTTGAGTCCATAAACCAGCTCAAGATCCAGAGGAGCTTC 216  
Qy 1363 ATGCTAAGTCTTCAGAGACCCCAAGGCTATGTCCAGACCTGCTCGCTCCAGAGC 1422  
Db 217 ATGCTAAGTCTTCAGAGACCCCAAGGCTATGTCCAGACCTGCTCGCTCCAGAGC 276  
Qy 1423 CGGACCTCAAGGTGATGACAGATGTAGCCGGCAACCTGAAAGAGAGCGCGGCTGAG 1482  
Db 277 CGGACCTCAAGGTGATGACAGATGTAGCCGGCAACCTGAAAGAGAGCGCGGCTGAG 336  
Qy 1483 TTCTACCAACAGCCCTGTGTCACAGAGCGCTGAGTGCCTACTTCTACTGCAAGATCCAG 1542

Db 337 TTTCACCACAGCGCTTCCAGGAGCGCTGAGTGTGCCTGCTACTTCTACTGCAAGATCCAG 396  
Qy 1543 CAGCGCAGGAGGAGCTGAGCAGTCTGCTGTTGCGCAACACCTAGGAGCCCAAAAC 1602  
Db 397 CAGCGCAGGAGGAGCTGAGCAGTCTGCTGTTGCGCAACACCTAGGAGCCCAAAAT 456  
Qy 1603 AAGCAG 1608  
Db 457 AAGCAG 462  
RESULT 3  
US-09-918-995-26465  
; Sequence 26465, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 26465  
; LENGTH: 474  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(474)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-26465

Query Match 23.9%; Score 412.2; DB 11; Length 474;  
Best Local Similarity 98.1%; Pred. No. 7.9e-109;  
Matches 417; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
Qy 793 CCTGACAAACACCTGCTTGTGATGCGATCGGACCCAGAGCCAGGAGAGCGGCTTC 852  
Db 50 CCTGACAAACACCTGCTTGTGATGCGATCGGACCCAGAGCCAGGAGAGCGGCTTC 109  
Qy 853 CAGGTGAAACCGGCTGGGACCTCAGTGTGCGTGACGCTGCTCCTCATCTGAGTAC 912  
Db 110 CAGGTGAAACCGGCTGGGACCTCAGTGTGCGTGACGCTGCTCCTCATCTGAGTAC 169  
Qy 913 CAGCCTCCCAAGTTCAAACCTGGATCCCGCTAGCCGCTGCTGGGCTGCACACACAG 972  
Db 170 CAGCCTCCCAAGTTCAAGATCGGATCCCGCTATCCCGCTGCTGGGCTGCACACACAG 229  
Qy 973 AGCGCTCAGCCATTTGCGAGGCTGCTGAGAGCATGTGTAAGACCAACAGGCTGCAGAC 1032  
Db 230 AGCGCTCAGCCATTTGCGAGGCTGCTGAGAGCATGTGTAAGACCAACAGGCTGCAGAC 289  
Qy 1033 TCCCATGACAGGAATACATCAATGGGACAAAGTATTTCCAGAGATTTTGTGTTGCTCC 1092  
Db 290 TCCCATGACAGGAATACATCAATGGGACAAAGTATTTCCAGAGATTTTGTGTTGCTCC 349  
Qy 1093 CGGCTGAAGTTTCTGAGATTTCCCGCTCAGCGCTCAGAGCTGCTATTGCCCTGACCCA 1152  
Db 350 CGGCTGAAGTTTCTGAGATTTACCCAGCGCTCAGAGCTGCTATTGCCCTGACCCA 409  
Qy 1153 ATTGCTATCAACCATGTATCAGCGTGGACCTTTACAGACGAGAGAGAGCGTGTAT 1212  
Db 410 ATTGCTATCAACCATGTATCAGCGTGGACCTTTACAGACGAGAGAGAGCGTGTAT 469  
Qy 1213 GACAT 1217  
Db 470 GACAT 474





Query Match 19.7%; Score 340.4; DB 11; Length 471;  
Best Local Similarity 96.8%; Pred. No. 4.4e-88;  
Matches 358; Conservative 0; Mismatches 11; Indels 1; Gaps 1;  
QY 1348 ATCCAGAGGACTTCATGCTTAAGCTTCTCCAGAGACCCCAAGGCTGTATGTCAGAGACCTG 1407  
DB 55 ATCCAGAGGAGCTTCATGCTTAAGCTTCTCCAGAGACCCCAATTTGCTATGTCAGAGATCTG 114  
QY 1408 CTCGCTCCAGAGCGGAGCTCAAGGTGATGACAGATGAGCGGCAACCCCTGAAGAG 1467  
DB 115 CTCGCTCCAGAGCGGAGCTCAAGGGATGACAGATGAGCGGCAACCCCTGAAGAG 174  
QY 1468 GAGCCCGGGCTGAGTTTACACAGAGCCCTGGTCCAGGAGCGGCTCAGTGGCTACTTTC 1527  
DB 175 GAGCCCGGGCTGAGATCTACACAGAGCCCTGGTCCAGGAGCGGCTCAGTGGCTACTTTC 234  
QY 1528 TACTCAAGATCCAGACGCGAGGAGCTGGAGCAGTGCCTGGTTGTGCGCAACACC 1587  
DB 235 TACTCAAGATCCAGACGCGAGGAGCTGGAGCAGTGCCTGGTTGTGCGCAACACC 294  
QY 1588 TAGGAGCCCAAAACAAGCAGACACGAGACTTTCAGCCGTGTCGCGGCGCCAGCATTT 1647  
DB 295 TAAGAGCCCAAAATAAGCAGACACGAGAACTTTCAGCCGTGTCGCGGCGCCAGCATTT 354  
QY 1648 TTGCCCCGGGCTCCAGC-TCACTCTCTGCCACCTTGGGGTGTGGGGTGGATTAAAAAGT 1706  
DB 355 TTGCCCCGGGCTCCAGCATCACTCTCTGCCACCTTGGGGTGTGGGGTGGATTAAAAAGT 414  
QY 1707 CATTCATCTG 1716  
DB 415 CATTCGCTCTG 424

## RESULT 7

US-09-960-352-12390  
; Sequence 12390, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengping  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; PRIOR FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 12390  
; LENGTH: 285  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 53-LIB3058-001-Q1-K1-F2  
US-09-960-352-12390

Query Match 12.0%; Score 206.2; DB 10; Length 285;  
Best Local Similarity 83.0%; Pred. No. 2.1e-49;  
Matches 235; Conservative 0; Mismatches 48; Indels 0; Gaps 0;  
QY 559 GTGACATCCAGGAGGCTCTGAAGAGGCCCATGAAGCAAAAGCGGAAGCTGCGACTCTAT 618  
DB 3 GTGGTCATGCTGATGCTCTGAGCTGTTACGAGCAGAGAGCGCAACTGGTCTGCTAT 62  
QY 619 ATCTCCAACTTTTAAACCTGCGAAGTCTGATGCTGAGGATTCGAGCGGAGCATTTGCC 678  
DB 63 ATCTCCAACTTTTAAACCTGCTAAGCCCGATGCTAGGATTTCTGAGCGGAGCATTTGCC 122  
QY 679 TCCTGGAGCTACGGGTGGAGGAGCTCCTGGATGATCCAGCAAAACAGAGCGGAAG 738  
DB 123 TCCTGGAGCTGCTGGTGGAGGAGCTCCTGGATGATCCAGCAAAACAGAGCGGAAG 182  
QY 739 TTCTCTCTTTTCTTCAAGAGTTTGGTCTATCGAGCTGGAGCAAAAGATCTTTTATGGCCCTGAC 798

DB 183 TTCTCGTCTTCAAGAGCTCGGTCAATTGATCTGACATAGACCTTTATGGCCCTGAC 242  
QY 799 AACCACTCTGTTAGTGGCATCGACACCCACGACCCAGGAGA 841  
DB 243 AACCACTGTTGGAGTGGCACCTGACACCCACCACTCAGAGA 285

## RESULT 8

US-09-879-536-796/C  
; Sequence 796, Application US/09879536  
; Patent No. US20020144298A1  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/879,536  
; CURRENT FILING DATE: 2001-09-21  
; PRIOR APPLICATION NUMBER: US 60/088,801  
; PRIOR FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 796  
; LENGTH: 452  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(452)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-879-536-796

Query Match 11.9%; Score 205; DB 10; Length 452;  
Best Local Similarity 73.4%; Pred. No. 5.6e-49;  
Matches 312; Conservative 0; Mismatches 90; Indels 23; Gaps 3;  
QY 480 GCTGGTCCCGAGTCCAGGCTTACATGGACCTCTTGGCATTTGAGAGGAACTGGATCA 539  
DB 424 GCAGGTACCAGAAATCCAGGCTTATATGGATCTCTTGGCTTTTGAAGG-AACTGGACCA 366  
QY 540 AACCATCATCGGAGGGGTGACATCCAGGAGGCTCTGAAGAGGCCATGAAGCAAAA 599  
DB 365 GACTATCATGAGAAC-GGCTAGATATCCAAGAGGCTTGAAGCGTCCCATCAAGCAAAA 307  
QY 600 GCGAAGCTCGACTCTATATCCCAACACTTTTAAACCCTGCGAAGTCTGATGCTGAGGA 659  
DB 306 ACGGAAGCTCGAATTTTCAATTTTCAACACTTTTCAATCCGGCTAAGTCAGATCCGAGGA 247  
QY 660 TTCCGAGCGGAGCATTTCCCTCCCTGGGAGCTACGGGTGGAGGGAGGCTCTCTGG----- 712  
DB 246 TGGGAAGGGACGGTGGCTTCTCGGAGCTTCGGGTAGAGGAGCGGCTCCTGGAGGATTC 187  
QY 713 -----ATGATCCAGCAAAACAGAGCGGAAGTTCCTCTCTTTTCAAGAG 758  
DB 186 AGCCTTGTCCAAATATGATGCCACTAAACAAAGAGGAAGTCTCTCTCTTTTAAATC 127  
QY 759 TTTGGTCATCGAGCTGACAAAGATCTTTATGGCCCTGCAACACCTCGTTGATGGCA 818  
DB 126 CTGGGTGATTGAATGGCAAGACCTGTATGGGCCAGAACCACTCTGTTAGTATGGCA 67  
QY 819 TCAGACACCCAGCAGCCAGGAGACGGGCTTCCAGGTGAAACGGGCTGGGACCTGAG 878

Db 66 CAGAGCCGACACTACCCAGAGACCATGCTTTCAGTGAACGGCCGGAGACCTGAA 7  
QY 879 TGTGC 883  
Db 6 TGTAC 2

RESULT 9  
US-09-918-995-2832

/ Sequence 2832, Application US/09918995  
/ Publication No. US20030073623A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Hyseq, Inc.  
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
/ FILE REFERENCE: 20411-756  
/ CURRENT APPLICATION NUMBER: US/09/918,995  
/ CURRENT FILING DATE: 2001-07-30  
/ PRIOR APPLICATION NUMBER: US/09/235,076  
/ PRIOR FILING DATE: 1999-01-20  
/ NUMBER OF SEQ ID NOS: 38054  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 2832  
/ LENGTH: 495  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (1)...(495)  
/ OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-2832

Query Match 11.8%; Score 203; DB 11; Length 495;  
Best Local Similarity 75.6%; Pred. No. 2.2e-48;  
Matches 251; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 662 CCGAGCGGACGATGCTCTCGGAGCTACGGGTGAGGGGAAAGCTCCGTGATGCCA 721  
Db 164 CAGGGGACAAAGGGGCTCTCGGAACTCGAGTGAAGAAAGCTGGATGATCTTA 223  
QY 722 GCAAAAGAGGAGAGTTCTCTTCTTCAAGATTGTGATGAGCTGAGCAAAAG 781  
Db 224 GCAAAAGAGGAGAGTTCTCTTCTTCAAGATTGTGAGCTGAGCAAAAG 283  
QY 782 ATCTTATGCGCTTGAACAACCTGTTGAGTGGATGAGACCCAGACCCAGAGAG 841  
Db 284 AGCTGTAAGGGCTGACATCACTGTGTGAGTGGACCGGATGCCACACCCAGAGAG 343  
QY 842 CGAGCGGCTTCCAGAGTGAAGAGCGCTGGGAGCTGAGTGGCTGACAGCTGCTCTTA 901  
Db 344 CAGATGCTTCCAAAGTAAACGGCTGGAGACCTCAACGTCACCTCTCTCTCA 403  
QY 902 TGTGACTACCAAGCTCCCAATTAAGTGGATCCCGCTAGCCGCTGCTGGAGC 961  
Db 404 TGTGATCATCATGAGCTCTCCCAATTAAGTGGATCCCGCTGAGAGGCTTGTGAG 463  
QY 962 TGCACACAGAGCGCTCAGCCATGTCAG 993  
Db 464 TGCACACAGAGCAAAAGCCGATCATGTCAG 495

## RESULT 10

US-09-918-995-34033  
/ Sequence 34033, Application US/09918995  
/ Publication No. US20030073623A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Hyseq, Inc.  
/ TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
/ FILE REFERENCE: 20411-756  
/ CURRENT APPLICATION NUMBER: US/09/918,995  
/ CURRENT FILING DATE: 2001-07-30

/ PRIOR APPLICATION NUMBER: US/09/235,076  
/ PRIOR FILING DATE: 1999-01-20  
/ NUMBER OF SEQ ID NOS: 38054  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 34033  
/ LENGTH: 420  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-09-918-995-34033

Query Match 10.5%; Score 180.4; DB 11; Length 420;  
Best Local Similarity 94.4%; Pred. No. 7e-42;  
Matches 187; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 903 GCTGACTACCAAGCTCCCAAGTTCAATGATCCCGGCTAGCCGCTGCTGGGCT 962  
Db 3 GCTGACTACCAAGCTCCCAAGTTCAATGATCCCGGCTAGCCGCTGCTGGGCT 62  
QY 963 GCACACAGAGCCGCTGAGCCATTTGTCAGGCGCTGTGAGAGTGTGAAGACCAACAG 1022  
Db 63 GCACACAGAGCCGCTGAGCCATTTGTCAGGCGCTGTGAGAGTGTGAAGACCAACAG 122  
QY 1023 GCTGAGAGCTCCATGACAAAGAAATACATCAATGAGGAGCAAGTATTTCCAGCATTT 1082  
Db 123 GCTGAGAGCTCCATGACAAAGAAATACATCAATGAGGAGCAAGTATTTCCAGCATTTCC 182  
QY 1083 TGATGTCCCGGCTGAA 1100  
Db 183 TGATGAGCAAGAGCAGTA 200

## RESULT 11

US-09-879-536-645/C  
/ Sequence 645, Application US/09879536  
/ Patent No. US20020144298A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Endege, Wilson O.  
/ APPLICANT: Steinmann, Kathleen E.  
/ APPLICANT: Astle, Jon H.  
/ APPLICANT: Burgess, Christopher C.  
/ APPLICANT: Bushnell, Steven E.  
/ APPLICANT: Carroll III, Eddie  
/ APPLICANT: Carino, Theodore J.  
/ APPLICANT: Derti, Adnan  
/ APPLICANT: Ford, Donna M.  
/ APPLICANT: Lewis, Marcia E.  
/ APPLICANT: Monahan, John E.  
/ APPLICANT: Schlegel, Robert  
/ TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
/ FILE REFERENCE: CCD-257 (US)  
/ CURRENT APPLICATION NUMBER: US/09/879,536  
/ PRIOR FILING DATE: 2001-09-21  
/ PRIOR APPLICATION NUMBER: US 60/088,801  
/ PRIOR FILING DATE: 1998-06-10  
/ NUMBER OF SEQ ID NOS: 850  
/ SOFTWARE: FastSeq for Windows Version 3.0  
/ SEQ ID NO 645  
/ LENGTH: 690  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc\_feature  
/ LOCATION: (1)...(690)  
/ OTHER INFORMATION: n = A,T,C or G  
US-09-879-536-645

Query Match 10.4%; Score 179.2; DB 10; Length 690;  
Best Local Similarity 64.5%; Pred. No. 1.8e-41;  
Matches 323; Conservative 0; Mismatches 169; Indels 9; Gaps 4;

QY 1112 TTCCGAGGCTCAGAGCCCTGCTATTGCCCCCTGACCAATGTCATCA-ACCATGTC 1170

```

RESULT 13
US-09-860-670-247/c
; Sequence 247, Application US/09860670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA127P1
; CURRENT APPLICATION NUMBER: US/09/860,670
; CURRENT FILING DATE: 2001-05-21
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 247
; LENGTH: 8788
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-860-670-247

Query Match          9.9%; Score 170.4; DB 10; Length 8788;
Best Local Similarity 98.9%; Pred. No. 1.5e-38;
Matches 182; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1534 AAGATCCAGCAGCGCAGGAGGAGCTGGAGCAGTTCAGCCGTGTCCTGGTGTGCGCAACACCTAGGAG 1593
      |||
Db 6674 AAGATCCAGCAGCGCAGGAGGAGCTGGAGCAGTTCAGCCGTGTCCTGGTGTGCGCAACACCTAGGAG 6615

QY 1594 CCCAAAAACAGCAGCAGCAGCAGCGAACTTTTCAGCCGTGTCCTGGGCCCCCAGCATTTTGGCCC 1653
      |||
Db 6614 CCCAAAAAATAGCAGCAGCAGCAGCGAACTTTTCAGCCGTGTCCTGGGCCCCCAGCATTTTGGCCC 6555

QY 1654 CGGGCTCCAGC-TCACCTCCTCTGCCACCTTGGGGTGTGGGGCTGGATTAAAGTCATTCA 1712
      |||
Db 6554 CGGGCTCCAGCATCATCTCTCTGCCACCTTGGGGTGTGGGGCTGGATTAAAGTCATTCA 6495

QY 1713 TCTG 1716
      |||
Db 6494 TCTG 6491

RESULT 14
US-09-918-995-27234
; Sequence 27234, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 27234
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(440)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27234

```

```

RESULT 14
US-09-918-995-27234
; Sequence 27234, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27234
; LENGTH: 440
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(440)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-27234

Query Match          9.4%;   Score 162;   DB 11;   Length 440;
Best Local Similarity 97.1%;   Pred. No. 15e-36;
Matches 165; Conservative 0; Mismatches 5; Indels 0; Gaps 0

Qy          910   TACAGCGCTCCCGAGTTCAACTGGATGCCCGCTAGCCCGGCTGTGTGGGGCTTGCAACA 969

```

Db 47 TGCACGCTCCCGAGTTCAACTGATCCCGCTAGTTGGCTGCTGGGCTGCACACA 106  
Qy 970 CAGAGCGGCTCAGCCATTGTTCAGGCGCTGTGGCAGTATGTAAGCAACAGGCTGCAG 1029  
Db 107 CAGAGCGGCTCAGCCATTGTTCAGGCGCTGTGGCAGTATGTAAGCAACAGGCTGCAG 166  
Qy 1030 GACTCCCATGACAGGATATCATCATGAGGAGCAAGTATTTCCAGCAGAT 1079  
Db 167 GACTCCCATGACAGGATATCATCATGAGGAGCAAGTATTTCCAGCAGAT 216

## RESULT 15

US-10-070-676-23/c  
Sequence 23, Application US/10070676  
Publication No. US20030059788A1  
GENERAL INFORMATION:  
APPLICANT: Toque, Bruno  
APPLICANT: Bracco, Laurent  
APPLICANT: Schweighoffer, Fabien  
TITLE OF INVENTION: Genetic Markers of Toxicity, Preparation  
TITLE OF INVENTION: and Uses  
FILE REFERENCE: 50146/003002  
CURRENT APPLICATION NUMBER: US/10/070,676  
CURRENT FILING DATE: 2002-03-06  
PRIOR APPLICATION NUMBER: PCT/FR00/02503  
PRIOR FILING DATE: 2000-09-12  
PRIOR APPLICATION NUMBER: FR 99/11405  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: 09/456,370  
PRIOR FILING DATE: 1999-12-08  
NUMBER OF SEQ ID NOS: 37  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 23  
LENGTH: 310  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 3, 33, 34  
OTHER INFORMATION: n = A,T,C or G  
US-10-070-676-23

Query Match 8.0%; Score 137.8; DB 14; Length 310;  
Best Local Similarity 69.1%; Pred. No. 1.3e-29;  
Matches 201; Conservative 0; Mismatches 89; Indels 1; Gaps 1;

Qy 991 CAGGCCCTGTGCAGTATGTAAGCAACAGGCTGCAGACTCCCATGACAAGGATATC 1050  
Db 306 CAGGCCCTGTGCTTACATCAAGCAACAGGCTGCAGATGGGACAGCGGAGTAC 247  
Qy 1051 ATCAATGGGACAGATATTTCCAGCAGATTTTGAATGTCCCGGCTGAAGTTTCTGAG 1110  
Db 246 ATCAACTGCAACGTTACTTCGCCAGATCTTCAGTTG-GGCCGACTCCGTTTCTCCGAG 188  
Qy 1111 ATTCCCGACGCTCAACGCCCTGCTATGCCCCCTGACCCCAATTTGTCATCAACCATGTC 1170  
Db 187 ATTCCCATGAAGCTGCGAGGTTGCTGCAGCATTCAGACCCCAATTTGTCATCAACCATGTC 128  
Qy 1171 ATCAGCGTGAACCTTCAGACCGAAGAGACAGCGTGTATGACATTTGACGTGAGGTG 1230  
Db 127 ATTAGTGTGACCTTAACGACGAGAGAGACAGCGTGTATGACATTCGATGTGAGGTG 68  
Qy 1231 GAGGAGCCATTAAAGGGGAGATGAGCAGCTTCTCTATTCAGGGCCAAAC 1281  
Db 67 GAGGAGCCACTGAAGGCCCAATGAGCAATTTTNNGGCTCTACCAACCATC 17

Search completed: October 11, 2003, 12:32:45  
Job time : 358 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 11, 2003, 09:10:51 ; Search time 2542 Seconds  
(without alignments)  
16483.445 Million cell updates/sec

Title: U66619  
Perfect score: 1724  
Sequence: 1 GAATTCGGCGCGAGCGGCC.....GTCAATCATCTGGGAATTC 1724

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: em\_esthum:\*  
18: em\_estinv:\*  
19: em\_estpin:\*  
20: em\_estvrt:\*  
21: em\_estfun:\*  
22: em\_estmam:\*  
23: em\_estmus:\*  
24: em\_estpro:\*  
25: em\_estrod:\*  
26: em\_estphg:\*  
27: em\_estvrt:\*  
28: gb\_gse1:\*  
29: gb\_gse2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1183.4	58.6	1669	AK005094	AK005094 Mus muscu
2	1016	58.9	1201	AL571327	AL571327 AL571327
3	965.6	56.0	1201	AL565228	AL565228 AL565228
4	796.8	46.2	1028	BM552001	BM552001 AGENCOURT

C	5	783.8	45.5	834	9	AL546132
C	6	761	44.1	906	13	BX369145
C	7	750.4	43.5	1050	12	BI524726
C	8	741.8	43.0	851	12	BI524142
C	9	740.6	43.0	904	13	BO720110
C	10	730.2	42.4	842	12	BI553787
C	11	728.8	42.3	807	12	BI520621
C	12	724.6	42.0	958	12	BI520614
C	13	698.4	40.5	1201	9	AL533695
C	14	694.4	40.3	837	14	CD517671
C	15	685.2	39.7	894	14	CD558853
C	16	681.2	39.5	3206	11	BC000063
C	17	679.6	39.4	1129	10	BE688178
C	18	674.8	39.1	947	13	BO955489
C	19	672.6	39.0	859	14	CA983665
C	20	672.4	39.0	710	12	BM979567
C	21	660.4	38.3	714	9	AI091285
C	22	659.8	38.3	819	12	BI906319
C	23	657.4	38.1	701	13	BU731539
C	24	639.6	37.1	1163	12	BM547645
C	25	634.2	36.8	704	13	BO954817
C	26	631.6	36.6	911	13	BO670467
C	27	624.6	36.2	674	9	AI233028
C	28	624.2	36.2	732	14	CA323812
C	29	616.6	35.8	805	11	AK019374
C	30	616.4	35.8	863	12	BI601560
C	31	615.6	35.7	815	12	BG825152
C	32	614.6	35.6	717	9	AA902273
C	33	612.6	35.5	811	14	BY712841
C	34	610.2	35.4	813	10	BF972004
C	35	607.8	35.3	729	12	BI598614
C	36	607	35.2	615	9	AI300927
C	37	596.8	34.6	805	10	BG170919
C	38	595.8	34.6	961	10	BG115124
C	39	595.2	34.5	877	13	BUS21158
C	40	592.2	34.4	936	13	BUS91019
C	41	586.8	34.0	947	13	BO940444
C	42	582.2	33.8	664	9	AI564026
C	43	581	33.7	628	14	CA449683
C	44	576.4	33.4	603	12	BG819745
C	45	569.8	33.1	679	10	BF722549

## ALIGNMENTS

RESULT 1  
LOCUS AK005094  
DEFINITION Mus musculus adult male cerebellum cDNA, RIKEN full-length enriched library, clone:15000010J14 product:SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3, full insert sequence.  
ACCESSION AK005094  
VERSION AK005094.1 GI:12836786  
KEYWORDS HTC, CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE  
1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL MEDLINE  
PUBMED 10349636  
AUTHORS  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL MEDLINE  
20499374





```

QY 636 CCTGCGAAGTCTGATGCTGAGGATTCCGACGCGACGATTCCTCTGGAGCTACGGGT 695
    |||||
DB 589 CCTCGGAAAGCTGATGCGGAAAGCTCTGATGGACACATTCCTCTGGAGAGTGGGGT 648
QY 696 GAGAGGGAAGCTCTGGATGATCCGAAACGAAAGCGAAATCTCTTCTTTCTTCAA 755
    |||||
DB 649 GAGAGGGAAGCTCTTGGATGATCTGATGAAGAGAAAGAAATTTCTTCTTCTTCAA 708
QY 756 GAGTTTGTCTGATGAGCTGAGCAGAAAGATTTTATGGCCCTGACACACCTCTGTTAGAG 815
    |||||
DB 709 GAGTTTGTCTGATGAGCTGAGCAGAAAGATTTTATGGCCGACAGACACCTCTGTTAGAG 768
QY 816 GCATCGACACCCAGCAGCAGGAGCGAGCGCTTCAGAGTAAACGGCTGGGAGCT 875
    |||||
DB 769 GCACCGGAAACCCAGCAACCCAGAAACAGATGGGTTCCAGTGAAGACACAGGAGACTT 828
QY 876 GAGTGTGCGCTGACGCTGTCTCTCATGTCTGATCAAGCCCTCCCACTTCAAACTGA 935
    |||||
DB 829 GAGTGTGCGCTGACGCTGTCTCTCATGTCTGATCAAGCCCTCCCACTTCAAACTGA 888
QY 936 TCCCCCCTAGCCCGCTGTCTGGGCTGCACACAGAGCCGCTGAGCCATTTGTCCAGGC 995
    |||||
DB 889 CCCCCCTTAGCCCGCTGTCTGGGCTGCACACAGAGCCGCTGAGCCATTTGTCCAGGC 948
QY 996 CCGTGGCAGTATGTGAAGACCAACAGGCTGCAGAGCTCCATGACAGAAATACATCAA 1055
    |||||
DB 949 ACTGTGGCAGTATGTGAAGACCAACAGGCTGCAGAGCTCCATGACAGAAATACATCAA 1008
QY 1056 TGGGGCAAGATATTTTCAGCAGATTTTGTATGTTCCCGGCTGAAGTTTCTGAGATTC 1115
    |||||
DB 1009 TGGCGCAAGATATTTTCAGCAGATTTTGTATGTTCCCGGCTGAAGTTTCTGAGATTC 1068
QY 1116 CCAAGCCTCTACAGCCCTGTCTATTTGCCCTGACCCCAATTGTCTAACCATCTCATCAG 1175
    |||||
DB 1069 CCAAGCCTCTACAGCCCTGTCTATTTGCCCTGACCCCAATTGTCTAACCATCTCATCAG 1128
QY 1176 CCGTGGCAGTATGTGAAGACCAACAGGCTGCAGAGCTCCATGACAGAAATACATCAA 1235
    |||||
DB 1129 CCGTGGCAGTATGTGAAGACCAACAGGCTGCAGAGCTCCATGACAGAAATACATCAA 1188
QY 1236 GCCATTAAAGGGGAGATGAGCAGCTTCTCTATTCACAGGCGCAACACAGAGAGATCAG 1295
    |||||
DB 1189 ACCGCTGAAGGGAGATGAGTATGCTTCTCTGTCACAGGCGCAACACAGAGAGATCAG 1248
QY 1296 TCCCTGTGACAGTATGATCCATGAGACGATTGATCCATTAACAGCTCAAGTCCAGAG 1355
    |||||
DB 1249 TCCCTGTGACAGTATGATCCATGAGACGATTGATCCATTAACAGCTCAAGTCCAGAG 1308
QY 1356 GGAATTGATGCTAAGCTTCTCCAGAGACCCCAAGGCTATGTCAGAGACTGCTCCGCTC 1415
    |||||
DB 1309 GGAATTGATGCTAAGCTTCTCCAGAGACCCCAAGGCTATGTCAGAGACTGCTCCGCTC 1368
QY 1416 CCAGAGCCCGGAGCCTCAAGGTGATGACAGATGATGAGCGGCAACCTGAAGAGAGCGCCG 1475
    |||||
DB 1369 CCAGAGCCCGGAGCCTCAAGGTGATGACAGATGATGAGCGGCAACCTGAAGAGAGCGCG 1428
QY 1476 GGCTGAGTTCTACACAGCCCTGTGTCAGAGAGGCGCTGAGTCTGCTTCTTCTGCAA 1535
    |||||
DB 1429 GGCTGAGTTCTACACAGCCCTGTGTCAGAGAGGCGCTGAGTCTGCTTCTTCTGCAA 1488
QY 1536 GATCCGAGAGCGGAGAGAGAGTGTGAGAGTGTGCTGTTGCGCAACCTGAGAGAGCG 1595
    |||||
DB 1489 GATCCGAGAGCGGAGAGAGTGTGAGAGTGTGCTGTTGCGCAACCTGAGAGAGCG 1548
QY 1596 CAAAAACAAGCAGCAGCAGAACTTCA 1624
    |||||
DB 1549 CGTGAACAAGCGTCAAGGTGAGACCA 1577
    |||||

```

RESULT 2  
AL571327/c  
LOCUS AL571327 1201 bp mRNA linear EST 31-MAY-2003

```

DEFINITION AL571327 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
ACCESSION AL571327
VERSION AL571327.2 GI:31292727
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 16, 2001 this sequence version replaced gi:12928512.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4359.r For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOD1009AE05NP1cluster=4359.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSOD1009AE05NP1.

FEATURES
    source
        1..1201
            location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSOD1009Y109"
                /tissue_type="PLACENTA COT 25-NORMALIZED"
                /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                /note="1st strand cDNA was primed with a NotI-oligo(dT)
                primer. Five prime end enriched, double-strand cDNA was
                digested with Not I and cloned into the Not I and EcoR V
                sites of the pCMVSPORT 6 vector. Library was normalized."
                sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 231 a 309 c 341 g 283 t 37 others
ORIGIN
Query Match 58.9%; Score 1016; DB 9; Length 1201;
Best Local Similarity 96.1%; Pred. No. 3,9e-222;
Matches 1097; Conservative 15; Mismatches 22; Indels 8; Gaps 7;
QY 528 GAACTGATCAAAACCATCATCGGAGGGGCTGACATCCAGAGGCTCTGAAGAGCC 587
    |||||
DB 1161 GGAATGATCAAAACCATCAK--CGAAGCGGTGAMMMVCCAGRRGCTCTGAAGAGGCC 1104
QY 588 CATGAAGCAAAAGCGGAGAGCTCGACTATATCTCCAACTTTTAACCTGCCAAGTC 647
    |||||
DB 1103 CATGAAGCAAAAGCGGAGAG-YCGCACTMTATTTCTCAACACTTTTAACCTT-CGAGGCC 1046
QY 648 TGATGCTGAGATTCGACGAGAGATTCCTCTGAGAGCTTACGAGGTGAGAGGAGAGCT 707
    |||||
DB 1045 TGATGCTGAGATTCGACGAGAGATTCCTCTGAGAGCTTACGAGGTGAGAGGAGAGCT 986
QY 708 CCTGATGATC-CAGCAAAAGAGCGGAGTCTCTTTCTTTCAAGAGTTTGCTCA 766
    |||||
DB 985 CCTGATGATCCCCAGCAAAAGAGCGGAGTCTCTTTCTTTCAAGAGTTTGCTCA 926
QY 767 TCGAGCT-GGAAGAAAGATTTTATGAGCCCTGACCAACACCTGTTGAGTGGATCGGACA 825
    |||||
DB 925 TCGAGCTGGGAAAGATTTTATGAGCCCTGACCAACCACTGTTGAGTGGATCGGACA 866
QY 826 CCAAGCAACCAAGAGAGAGCGAGCTTCCAGTGTGAAGCGGCTGGGAGCTGAGTGTGCGC 885
    |||||
DB 865 CCAAGCAACCAAGAGAGAGCGAGCTTCCAGTGTGAAGCGGCTGGGAGCTGAGTGTGCGC 807
QY 886 TCGAGCTGCTCTCATGCTGAGTACCAAGCTTCCCAAGTTCAAACTGATCCCGCCTTA 945
    |||||
DB 806 TCGAGCTGCTCTCATGCTGAGTACCAAGCTTCCCAAGTTCAAACTGATCCCGCCTTA 747
    |||||

```

```

Qy 946 GCCCGGCTGCTGGGCTGCACACAGAGCGCTCAGCCATTGTCCAGGCGCTGTGGCAG 1005
Db 746 GCCCGGCTGCTGGGCTGCACACAGAGCGCTCAGCCATTGTCCAGGCGCTGTGGCAG 687
Qy 1006 TATGTGAAGACCAACAGAGCTCGAGACTCCCATGACAGGAATACATCAATGGGACAAG 1065
Db 686 TATGTGAAGACCAACAGAGCTCGAGACTCCCATGACAGGAATACATCAATGGGACAAG 627
Qy 1066 TATTTCCAGCAGATTTTGTGTTGCTCCCGGCTGAAGTTTCTTGAGATTCCCGAGCGCTC 1125
Db 626 TATTTCCAGCAGATTTTGTGTTGCTCCCGGCTGAAGTTTCTTGAGATTCCCGAGCGCTC 567
Qy 1126 ACAGCCCTGTATTGCCCCCTGACCAATTTGTATCAACACCATGTATCAGCGTGGACCCCT 1185
Db 566 ACAGCCCTGTATTGCCCCCTGACCAATTTGTATCAACACCATGTATCAGCGTGGACCCCT 507
Qy 1186 TCAGACCAAGAAGACAGCGCTGTATGACATTGACGTGGAGGTGGAGGAGCCATTAAAG 1245
Db 506 TCAGACCAAGAAGACAGCGCTGTATGACATTGACGTGGAGGTGGAGGAGCCATTAAAG 447
Qy 1246 GGGCAGATGACAGCTTCTCTATCCACGGCCCAACACAGCAGGAGATCAGTCTCTTGAC 1305
Db 446 GGGCAGATGACAGCTTCTCTATCCACGGCCCAACACAGCAGGAGATCAGTCTCTTGAC 387
Qy 1306 AGTAAGATCCATGACAGATTTAGTTCATATAACACAGCTCAAGATCCAGAGGATTTCAATG 1365
Db 386 AGTAAGATCCATGACAGATTTAGTTCATATAACACAGCTCAAGATCCAGAGGATTTCAATG 327
Qy 1366 CTAAGCTTCTCAGAGACCCCAAGGCTATGTTCAAGACCTGCTCCGCTCCAGAGCGG 1425
Db 326 CTAAGCTTCTCAGAGACCCCAAGGCTATGTTCAAGACCT-CTCCGCTCCAGAGCGG 268
Qy 1426 GACCTCAAGGTGATGACAGATTTAGCCGGCAACCTGAAAGAGGAGCGCGGCTGAGTTC 1485
Db 267 GACCTCAAGGTGATGACAGATTTAGCCGGCAACCTGAAAGAGGAGCGCGGCTGAGTTC 208
Qy 1486 TACCACCAAGCCCTGCTCCAGAGGCGCTCAGTGCCTACTTCTACTGCAAGATCCAGCAG 1545
Db 207 TACCACCAAGCCCTGCTCCAGAGGCGCTCAGTGCCTACTTCTACTGCAAGATCCAGCAG 148
Qy 1546 GCGAGGAGGAGCTGGAGCAGTGCCTGTTGTGGCAACACCTAGGAGCCCAAAAACAG 1605
Db 147 GCGAGGAGGAGCTGGAGCAGTGCCTGTTGTGGCAACACCTAGGAGCCCAAAAACAG 88
Qy 1606 CAGCAGCAGGAACCTTTCAGCGCTGCTCCGGGCGCCACGATTTTGGCCCGGCTCCAGCT 1665
Db 87 CAGCAGCAGGAACCTTTCAGCGCTGCTCCGGGCGCCACGATTTTGGCCCGGCGCCANCA 28
Qy 1666 CA 1667
Db 27 NA 26

```

```

RESULT 3
AL565228/c 1201 bp mRNA linear EST 12-MAY-2003
LOCUS CS0DF002YC19 3-PRIME, mRNA sequence.
DEFINITION AL565228 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
ACCESSION AL565228
VERSION AL565228.2 GI:30549059
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 16, 2001 this sequence version replaced gi:12916394.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

```

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 4359.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DF002AB10NP1&cluster=4359.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope location ID : CS0DF002AB10NP1.

#### FEATURES

source 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DF002YC19"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."  
 BASE COUNT 241 a 305 c 329 g 280 t 46 others  
 ORIGIN

Query Match 56.0%; Score 965.6; DB 9; Length 1201;  
 Best Local Similarity 93.5%; Pred. No. 1.4e-210;  
 Matches 1028; Conservative 16; Mismatches 50; Indels 5; Gaps 4;  
 Qy 595 CAAAGCGGAGCTGCGACTCTATATCTCAACACTTTTAACTCGGAAGTCTGATGCT 654  
 Db 1099 CATAAGCAAAARSCAARSTKCRCTTATTYCAACACTTTTACCCTCGGAASCTTTRTKCT 1040  
 Qy 655 GAGATTTCCGACGCGCAGCATGCTCTCTGGAGCTACGGTGGAGGGGAAGCTCTTGAT 714  
 Db 1039 GAGATTTCCGACGCGCAGCATGCTCTCTGGAGCTACGGTGGAGGGGAAGCTCTTGAT 980  
 Qy 715 GATCCAGCAAAACAGAACGGAAGTCTCTTCTTCAAGAGTTTGGTTCATCGAGCTG 774  
 Db 979 GATCCAGCAAAACAGAACGGAAGTCTCTTCTTCAAGAGTTTGGTTCATCGAGCTG 920  
 Qy 775 GACAAAGATCTTATGGCCCTGACAAACACCTCGTGAAGTGGCATCGGACCCACGACC 834  
 Db 919 GACAAAGATCTTATGGCCCTGACAAACACCTCGTGAAGTGGCATCGGACCCACGACC 860  
 Qy 835 CAGGAGACGCGCTTCCAGGTGAACGCGCTGGGACCTGAGTGTGCGCTGACCGCTG 894  
 Db 859 CAGGAGACGCGCTTCCAGGTGAACGCGCTGGGACCTGAAAGTGTGCTGCACGCTG 800  
 Qy 895 CTCCTCATGTGGACTACACAGCCTCCCGAGTTCAAACCTGGATCCCGCTAGCCCGGCTG 954  
 Db 799 CTCCTCATGTGGACTACACAGCCTCCCGAGTTCAAACCTGGATCCCGCTAGCCCGGCTG 740  
 Qy 955 CTGGGCTGCACACACAGAGCCGCTCAGCCATTGTTCAGGCCCTGTGGCAGTATGTGAAG 1014  
 Db 739 CTGGGCTGCACACACAGAGCCGCTCAGCCATTGTTCAGGCCCTGTGGCAGTATGTGAAG 680  
 Qy 1015 ACCACAGGCTGCAGGA-CTCCCATGACAGGAATACATCAATGGGACAGATTTTCCA 1073  
 Db 679 ACCACAGGCTGCAGATATCTCCCATGACAGGAATACATCAATGGGACAGATTTTCCA 620  
 Qy 1074 GCAGATTTTGTATGTCCTCCCGCTGAAGTTTCTGAGATTCCCAGCGCTCCACAGCCCT 1133  
 Db 619 GCAGATTTTGTATGTCCTCCCGCTGAAGTTTCTGAGATTCCCAGCGCTCCACAGCCCT 560  
 Qy 1134 GCTATTCGCCCTGACCCCAATTGTTCATCAACCATGTTCATGCGTGGACCCCTTCAGACCA 1193  
 Db 559 GCTATTCGCCCTGACCCCAATTGTTCATCAACCATGTTCATGCGTGGACCCCTTCAGACCA 500  
 Qy 1194 GAAGAAGACAGCGTGTCTATGACATTGACGTTGAGGTGGAGGAGCCATTAAAGGGGCGAGAT 1253  
 Db 499 GAAGAAGACAGCGTGTCTATGACATTGACGTTGAGGTGGAGGAGCCATTAAAGGGGCGAGAT 440

Qy	1254	GAGCAGCTTCCTCCATCAAGGCCAACAGAGAGATCAAGTCTCTGACAGTAAGAT	1313
Db	439	GAGCAGCTTCCTCCATCAAGGCCAACAGAGAGATCAAGTCTCTGACAGTAAGAT	380
Qy	1314	CCATGAGACGATYGATGCATPAAACAGACTCAAGATCCAGAGGAGACTTCATGTAAGCTT	1373
Db	379	CCATGAGACGATYGATGCATPAAACAGACTCAAGATCCAGAGGAGACTTCATGTAAGYTT	320
Qy	1374	CTCCAGAGACCCCAAAAGGCTATGTCCTCAAGACTGTCGCGCTCCAGAGCCGGGACCTCAA	1433
Db	319	CTCCAGAGACCCCAAAAGGYATATGTCCAAGACTG-TCCGCTCCAGAGCCGGGACCTCAA	261
Qy	1434	GGTATGACACATGTAGTACCGGCAACCTGGAAGAGAGCGCGGGCTGAGTTCTACCA	1493
Db	260	GGTATGACACATGTATGACCGGCAACCTGGAAGAGAGCGCGGGYTGAGTTCTACCA	201
Qy	1494	GCCCTGATCCCAAGAGGCCGTGATGCTACTTCTACTGCAAGATCCAGCAGCGCAGCA	1553
Db	200	GCCCTGATCCCAAGAGGCCGTGATGCTACTTCTACTGCAAGATCCAGCAGCGCAGCA	141
Qy	1554	GGAGCTGAGACGTGCGCTG-GTGTGAGCCAAACCTAGAGAGCCCAAAAACAAGCAGCAG	1612
Db	140	GGAGCTGAGACGTGCGCTGT	81
Qy	1613	ACGGAACCTTCAGCCGTGTCCCGGGGCCCAAGCATTTTGGC--CCGGGCTCAGCTCACTC	1670
Db	80	ACGGAACCTTCAGCCGTGTCCCGGGGCCCAAGCATTTTGGCCTCCGGGANNCAAAACACTC	21
Qy	1671	CTCTGCCACCTTTGGGCTGT 1689	
Db	20	CTCTGCCACCTTTGGGCTGT 2	

```
RESULT 4  
BMS52001 LOCUS.  
DEFINITION BMS52001 1028 bp mRNA EST-20-FEB-2002  
AGNCOURT_6543453 NIH_MGC_71 Homo sapiens CDNA clone IMAGE:5549450  
5' , mRNA sequence.  
ACCESSION BMS52001  
VERSION BMS52001.GI:18789498  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1028)  
NIH-MGC http://mgc.nci.nih.gov/.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgabbs@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
http://image.llnl.gov  
Plate: LHAM1260 row: a column: 03  
High quality sequence stop: 640.  
FEATURES  
source 1..1028 Location/Qualifiers
```

```

FEATURES
    source
        location/Qualifiers
            1..1028
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:5549450"
                /rissue_type="leiomyosarcoma"
                /lab_host="DH10B (phage-resistant)"
                /clone_1lb="NIH MG-71"
                /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt Average insert size 2.1 kb."

```

ORIGIN	BASE COUNT	251 a	288 c	282 g	205 t	2 others
Query Match	46.2%;	Score 796.8;	DB 12;	Length 1028;		
Best Local Similarity	98.7%;	Pred. No. 6.3e-112;				
Matches 834;	Conservative % 0;	Mismatches 8;	Indels 3;	Gaps 3;		
OY	441	GAAGATGGCTGCACAAAATCCCTCCCTCAAAAGATTCGGAGAGCTGGTCCCCGAGTCCACGGC	500			
Db	1	GAAGATGGCTGCACAAAATCCCTCCCTCAAAAGATTCGGAGAGCTGGTCCCCGAGTCCACGGC	60			
OY	501	TTACATGGAGCCTCTTGCAATTTGAGAGGAAACTGGATCAAAACATCATGCGGAAGGGGT	560			
Db	61	TTACATGGAGCCTCTTGCAATTTGAGAGGAAACTGGATCAAAACATCATGCGGAAGGGGT	120			
OY	561	GGACATCCAGAGGCTCTGAAGAGAGGCCATGAAACAAAAGCGAAGCTGCGACTCTATAT	620			
Db	121	GGACATCCAGAGGCTCTGAAGAGAGGCCATGAAACAAAAGCGAAGCTGCGACTCTATAT	180			
OY	621	CTCCACAACCTTTAAACCTCGGAAGTGTGATGCTGAGAGATTTCCGACGGCAGACTTGGCTC	680			
Db	181	CTCCACAACCTTTAAACCTCGGAAGTGTGATGCTGAGAGATTTCCGACGGCAGACTTGGCTC	240			
OY	681	CTGGAGAGTACGGGTGGAGGGGGAAGCTCCTGGATGATCCGACGAAACGAAGCGGAAGTT	740			
Db	241	CTGGAGAGTACGGGTGGAGGGGGAAGCTCCTGGATGATCCGACGAAACGAAGCGGAAGTT	300			
OY	741	CTCTTCCTTTCTTCAAGAGTTTGGTGCATCGAGCTGGAACAAAGATCTTTATGGCCCTGACAA	800			
Db	301	CTCTTCCTTTCTTCAAGAGTTTGGTGCATCGAGCTGGAACAAAGATCTTTATGGCCCTGACAA	360			
OY	801	CCACCTCGTGTGAGTGGCATCGGACACCCACGACCCAGAGAGCGGACTTCCAGGTGAA	860			
Db	361	CCACCTCGTGTGAGTGGCATCGGACACCCACGACCCAGAGAGCGGACTTCCAGGTGAA	420			
OY	861	ACGGGCTGGGGGACCTGAGTGTGGGCTGCACGCTGCTCTATGCTGGAAGTACACAGCGCTCC	920			
Db	421	ACGGGCTGGGGGACCTGAGTGTGGGCTGCACGCTGCTCTATGCTGGAAGTACACAGCGCTCC	480			
OY	921	CCAGTTCAAACTGATCCCGGCTGAGCCGCGCTGCGGCTGACACACAGAGCCGCTC	980			
Db	481	CCAGTTCAAACTGATCCCGGCTGAGCCGCGCTGCGGCTGACACACAGAGCCGCTC	540			
OY	981	AGCCATTGTCCAGGCGCTGTGGCGAGTATGTGAAGACCAACAGGCTGACAGACTCCCATGA	1040			
Db	541	AGCCATTGTCCAGGCGCTGTGGCGAGTATGTGAAGACCAACAGGCTGACAGACTCCCATGA	600			
OY	1041	CAAGGAATACATCAATGGGGACAAGTATTTTCAGACATTTTGTATTTGTCCTCCCGGCTGAA	1100			
Db	601	CAAGGAATACATCAATGGGGACAAGTATTTTCAGACATTTTGTATTTGTCCTCCCGGCTGAA	660			
OY	1101	GTTTTTCGAGATTCCTCCAGGCGCTGCACAGGCGCTGCTATTTGGCCCGCTGACCCAAATTGCAT	1160			
Db	661	GTTTTTCGAGATTCCTCCAGGCGCTGCACAGGCGCTGCTATTTGGCCCGCTGACCCAAATTGCAT	720			
OY	1161	CAACCATGTCAATCAGGCTGGAACCTTTCAGACCAAGAAAGACAGCGTGTATGACATTTGA	1220			
Db	721	CAACCATGTCAATCAGGCTGGAACCTTTCAGACCAAGAAAGACAGCGTGTATGACATTTGA	780			
OY	1221	CGTGG-AGGTGGAGAG-CCATTAAAGGGGCAATGACA-GCTTCTCTCTATCCAGGCG	1277			
Db	781	CGTGGAGGTGGAGAGAGCCATTAAAGGGGCAATGAGCAGGCTTCTCTCTATTCACGCG	840			
OY	1278	CAACC 1282				
Db	841	NCAC 845				
RESULT 5	AL546132/c	834 bp	mrna	linear	EST 31-MAY-2003	
LOCUS	AL546132	Homo sapiens	PLACENTA	COT 25-NORMALIZED	Homo sapiens	CDNA
DEFINITION	clone CS0D1025Y102 3-PRIME, mRNA sequence.					

Best Local Search  
Matches 8422

```

QY      399  GGAGCCACCGCCCGCGGAGCCGACAGTGCACAAGAGGAGGAAGATGGCTGACAAAT 458
        1  GGTGCCACCGCCCGCGGAGCCGACAGTGCACAAGAGGAGGAAGATGGCTGACAAAT 60
QY      459  CCTCCCTCAAGAGATTGCGGAGCTGTGTCCTCCGAGCTCCAGGCTTACATGGAACCTCTTGGC 518
        61  CTTCCCTCAAGAGATTGCGGAGCTGTGTCCTCCGAGCTCCAGGCTTACATGGAACCTCTTGGC 119
QY      519  ATTTGAGAGAACTGATCAAAACATCATGCGGAAAGGGGGTGGACATCCAGAGGCTCT 578
        120  ATTTGAGAGAACTGATCAAAACATCATGCGGAAAGGGGGTGGACATCCAGAGGCTCT 179
QY      579  GAAGAGGCCCATGAAAGGAGGAGGAGGAGCTGACTATATCTCCAAACCTTTTAAACC 638
        180  GAAGAGGCCCATGAAAGGAGGAGGAGGAGCTGACTATATCTCCAAACCTTTTAAACC 239
QY      639  TCGAAGTCTGATGCTGAGGATTCGACGCGACACATTCCTCTGGAGGCTACGGGTGA 698
        240  TCGAAGTCTGATGCTGAGGATTCGACGCGACACATTCCTCTGGAGGCTACGGGTGA 299
QY      699  GGGGAAGCTCTGATGATCCACGAAACAGAAAGCGAAGTTCTTTCTTCAAGAG 758
        300  GGGGAAGCTCTGATGATCCACGAAACAGAAAGCGAAGTTCTTTCTTCAAGAG 359
QY      759  TTTGTCATCGACCTGACAAAGATCTTTATGGCCCTGACAAACACCTGCTGAGTGA 818
        360  TTTGTCATCGACCTGACAAAGATCTTTATGGCCCTGACAAACACCTGCTGAGTGA 419
QY      819  TCGAAGACCCACGACCCAGAGAGAGAGCGGCTTCCAGGTAAAGCGCTGGGAGCTGAG 878
        420  TCGAAGACCCACGACCCAGAGAGAGAGCGGCTTCCAGGTAAAGCGCTGGGAGCTGAG 479
QY      879  TGTGCGCTGACGCTGCTCTCATGTGTGACTACAGGCTCCCAAGTTCAACTGATTC 938
        480  TGTGCGCTGACGCTGCTCTCATGTGTGACTACAGGCTCCCAAGTTCAACTGATTC 539
QY      939  CCGGCTAAGCCCGGCTGCTGGGGGTGACACACAGAGCCGCTGACCATTTGCCAGGCT 998
        540  CCGGCTAAGCCCGGCTGCTGGGGGTGACACACAGAGCCGCTGACCATTTGCCAGGCT 599
QY      999  GTGGCAGTATGTGAAGACCAACAGCTGACAGACTCCCATGACAAAGATACATCATG 1058
        600  GTGGCAGTATGTGAAGACCAACAGCTGACAGACTCCCATGACAAAGATACATCATG 658
QY      1059  GGAACAATTTTTCAGACAGATTTTGTATGTCTCCGCGTGAAGTTTCTGAGATCCCA 1118
        659  GGAACAATTTTTCAGACAGATTTTGTATGTCTCCGCGTGAAGTTTCTGAGATCCCA 718
QY      1119  GCGGCTCAC-AGGCTGCTATTTGCCCCCTGACCCCAATTTGTCATCAACCATGTATCAGG 1177
        719  GAGCCTTAAACAGGCTGCTATTTTGGCCCCCTGACCCCAATTTGTCATCAACCATGTATCAGG 778
QY      1178  TGGACCTTCAAGCAGAGAGAGAGACGCGTATGATGACATGACGAGGAGTGG--AGGA 1235
        779  GGGACCTTAAACAGGCTGCTATTTTGGCCCCCTGACCCCAATTTTGGAGAGGAGGAGGGA 838
QY      1236  GCCATTAAAGGGGAGATGAGCAGCTTCTCTTATCCAGCGCCACACAGCAGAGATCAG 1295
        839  GCCATTAAAGGGGAGATGAGCAGCTTCTCTTATCCAGCGCCACACATGAGGAGATCAGC 898
QY      1296  TCTCTTGG 1303
        899  TGTCTTGG 906

```

```

RESULT 7
BIS24726/c 1050 bp mRNA linear EST 29-AUG-2001
LOCUS 603051552T1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201316 3'
DEFINITION mRNA sequence.
ACCESSION BIS24726
VERSION BIS24726.1 GI:15349518

```

```

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1050)
AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1504 row: 9 column: 13
High quality sequence start: 2
High quality sequence stop: 832.
Location/Qualifiers
1..1050
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5201316"
/lab_host="DH10B"
/clone_1b="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: PCMV-SPOrt6;
Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleen. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
BASE COUNT 216 a 283 c 316 g 235 t
ORIGIN
Query Match 43.5%; Score 750.4; DB 12; Length 1050;
Best Local Similarity 94.6%; Pred. No. 2.7e-161;
Matches 876; Conservative 0; Mismatches 36; Indels 14; Gaps 9;
QY 747 TTTCTCAAGAGTTTGTGTCATGAGCTGACAAAGATCTTTATGGCCCTGACAAACACC- 805
        934 TTTTTCAGAGTTTGTGTCATGAGCTGACAAAGATCTTTATGGCCCTGACAAACACC- 875
QY 806 -TCGTTGAGTGCATGACGACCCACGA-----CCAGAGACGAGCG-CTTCCAGGTG 858
        874 TTGTTGAGTGCATTCGAGGACACCTCAGACCCCAAGAGAGAGCGAGCGCTTCCAGGTG 815
QY 859 AAAAGGCTCGGGAGACCTGAGT--GTGGCTGACAGCTGCTCATGCTGAGCTACACAGCC 917
        814 AAAAGGCTCGGGAGACCTGAGTGTGTGCTGACAGCTGCTCATGCTGAGCTACACAGCT 755
QY 918 TCC-CCAGTTCAAACTGAT--CCCGGCTAGCCCGGCT--GCTGGGGCTCAGACACAGAG 974
        754 TCTCTAGTTCAAACTGATTCCTCCGCTTAGCCCGCTTAGCCCGCTGCTGAGCTCAGACAG 695
QY 975 CCGCTCAGCATTGTCAGGCGCTGTGGCAGTATGTGAAGACCAACAGGCTCAGG-ACT 1033
        694 CCGCTCAGCATTGTCAGGCGCTGTGGCAGTATGTGAAGACCAACAGGCTCAGGTA 635
QY 1034 CCCATGACAAGAA--TACATCATGAGGAGCAAGTATTTCCAGAGATTTTGAATGTCC 1092
        634 CCCATGACAAGAAATTCATCATGAGGAGCAAGTATTTCCAGAGATTTTGAATGTCC 575
QY 1093 CGGCTGAAGTTTCTGAGATTCCTCCAGCGCTCAGACGCTCTGATATGCCCCCTGACCA 1152
        574 CGGCTGAAGTTTCTGAGATTCCTCCAGCGCTCAGACGCTCTGATATGCCCCCTGACCA 515

```

```
Qy 1153 ATTGTCATCAACCATGTTCATCAGCGTGACCCCTTCAGACCAGAAAGACACGCGTGTAT 1212
Db 514 ATTGTCATCAACCATGTTCATCAGCGTGACCCCTTCAGACCAGAAAGACGCGTGTAT 455
Qy 1213 GACATTGACGTGGAGGTGGAGAGCCATTAAAGGGGAGATGAGCAGGTTCTCTATCC 1272
Db 454 GACATTGACGTGGAGGTGGAGAGCCATTAAAGGGGAGATGAGCAGGTTCTCTATCC 395
Qy 1273 ACGGCCAACACGACGAGATCAGTCTCTGACACAGTAAAGTCCATGACAGCATTGAGTCC 1332
Db 394 ACGGCCAACACGACGAGATCAGTCTCTGACACAGTAAAGTCCATGACAGCATTGAGTCC 335
Qy 1333 ATAAACACAGCTCAAGATCCAGAGGAGCTTCATGCTTAAGCTTCTCCAGAGACCCCAAGGC 1392
Db 334 ATAAACACAGCTCAAGATCCAGAGGAGCTTCATGCTTAAGCTTCTCCAGAGACCCCAAGGC 275
Qy 1393 TATGTCCAAAGCTGCTCCGCTCCAGAGCCGGGACCTCAAGGTGATGACAGTGTAGCC 1452
Db 274 TATGTCCAAAGCTGCTCCGCTCCAGAGCCGGGACCTCAAGGTGATGACAGTGTAGCC 215
Qy 1453 GGCAACCTGAAGAGGAGCGGGCTGAGTTCTACACAGCCCTGCTGTCACGAGGCC 1512
Db 214 GGCAACCTGAAGAGGAGCGGGCTGAGTTCTACACAGCCCTGCTGTCACGAGGCC 155
Qy 1513 GTGAGTCTGCTACTTCTACTGCAAGATCCAGAGCGGCGAGGAGCTGGAGCAGTCTGCTG 1572
Db 154 GTGAGTCTGCTACTTCTACTGCAAGATCCAGAGCGGCGAGGAGCTGGAGCAGTCTGCTG 95
Qy 1573 GTTGTGGCCAAACACTAGAGGCCCAAAACAAAGCAGCAGCAGGAACTTTTCAGCGGTGC 1632
Db 94 GTTGTGGCCAAACACTAGAGGCCCAAAACAAAGCAGCAGCAGGAACTTTTCAGCGGTGC 35
Qy 1633 CCGGGCCCCAGCATTTTCCCGCGGC 1658
Db 34 CCGGGCCCCAGCATATAACCCCGGTC 9

RESULT 8
BI524142
LOCUS 603051552F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201316 5',
DEFINITION mRNA sequence.
ACCESSION BI524142
VERSION BI524142.1 GI:15348934
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11504 row: 9 column: 13
High quality sequence stop: 850.
Location/Qualifiers
1..851
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5201316"
/lab_host="DH10B"
/clone_lib="NIH_MGC_122"
```

/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;  
Site 1: NotI; Site 2: EcoRV (destroyed); RNA source  
anonymous pool of 24 week female lung, 16 week female  
spleen, and 20-22 week male spleens. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.4 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 026. Note:  
this is a NIH MGC Library."

BASE COUNT 208 a 243 c 231 g 168 t  
ORIGIN

Query Match 43.0%; Score 741.8; DB 12; Length 851;  
Best Local Similarity 95.6%; Pred. No. 2.4e-159;  
Matches 816; Conservative 0; Mismatches 33; Indels 5; Gaps 5;

Qy 652 GCTGAGGATTCGACGCGCAGATTGCTCTCTGGG-AGCTAGGGTGGAGGGAAGCTCT 710  
Db 1 GCTGAGGATTCGACGCGCAGATTGCTCTCTGGGAGCTACGGGTGGAGGGAAGCTCT 60  
Qy 711 GGATGATCCAGCAAAACAGAGCGGAGTTCTCTTTCTTCAAGAGTTTGGTCATCGA 770  
Db 61 GGATGATCCAGCAAAACAGAGCGGAGTTCTCTTTCTTCAAGAGTTTGGTCATCGA 120  
Qy 771 GCTGGACAAAGATCTTTATGGCCCTGACAAACCACTCGTTGAGTGGCATCGGACACCCAC 830  
Db 121 GCTGGACAAAGATCTTTATGGCCCTGACAAACCACTCGTTGAGTGGCATCGGACACCCAC 180  
Qy 831 GACCCAGGAGCGGCTTCAGAGTGAACCGCGCTGGGGACCTGAGTGTGCGCTGAC 890  
Db 181 GACCCAGGAGCGGCTTCAGAGTGAACCGCGCTGGGGACCTGAGTGTGCGCTGAC 240  
Qy 891 GCTGCTCCTCATGCTGAGCTACCGAGCTCCCGAGTTCAAACTGGATCCCGCTAGCCCG 950  
Db 241 GCTGCTCCTCATGCTGAGCTACCGAGCTCCCGAGTTCAAACTGGATCCCGCTAGCCCG 300  
Qy 951 GCTGCTGGGGCTGCACACACAGAGCGCTCAGCCATTTGTCAGGCCCTGTGGCAGTATGT 1010  
Db 301 GCTGCTGGGGCTGCACACACAGAGCGCTCAGCCATTTGTCAGGCCCTGTGGCAGTATGT 360  
Qy 1011 GAAGACCAACAGGCTGAGGACTCCCATATGACAGGAATATCATTAATGGGACAAAGTATTT 1070  
Db 361 GAAGACCAACAGGCTGAGGACTCCCATATGACAGGAATATCATTAATGGGACAAAGTATTT 420  
Qy 1071 CCAGCAGATTTTGAATGTCCTCCGGCTGAAGTTTCTGAGATTCCCGAGCGCTCAGAGC 1130  
Db 421 CCAGCAGATTTTGAATGTCCTCCGGCTGAAGTTTCTGAGATTCCCGAGCGCTCAGAGC 480  
Qy 1131 CCTGCTATTGCCCCCTGACCCCAATTGTTCATCAACCATGTTCATGCGTGGACCTTCAGAA 1190  
Db 481 CCTGCTATTGCCCCCTGACCCCAATTGTTCATCAACCATGTTCATGCGTGGACCTTCAGAA 540  
Qy 1191 CCAGAAAGACAGCGCTGCTATGACATTTGACGTTGAGGAGGAGGAGCCATTAAGGGGCA 1250  
Db 541 CCAGAAAGACAGCGCTGCTATGACATTTGACGTTGAGGAGGAGGAGCCATTAAGGGGCA 600  
Qy 1251 GATGAGCAGCTTCTCTATTCACGGCCAAACAGCAGGAGATCAGTCTCTGGACAGTAA 1310  
Db 601 GATGAGCAGCTTCTCTATTCACGGCCAAACAGCAGGAGATCAGTCTCTGGACAGTAA 660  
Qy 1311 GATCCATGAGACGA-TTGAGTCCATAAACCAAGCTCAAGATCCAGAGGAGCTTCATGTGTA 1369  
Db 661 GATCCATGAGACGA-TTGAGTCCATAAACCAAGCTCAAGATCCAGAGGAGCTTCATGTGTA 720  
Qy 1370 GCTTCTCCAGAGACCCCAAGGCTATGTCGAAGACCTGCTCCGCTCCAGAGCGGGACC 1429  
Db 721 AGCTTCTCCAGAGACCCCAAGGCTATGTCGAAGACCTGCTCCGCTCCAGAGCGGGAC-C 778  
Qy 1430 TCAGAGGTGATGACAGATGTAGCGGCAACCTGAAGAGGAGCGCGGCTGAGTTCTACC 1489  
Db 779 TCAGAGGTGATGACAGATGTAGCGGCAACCTGAAGAGGAGCGCGGG-TGAGTTCTAAC 837

QY	1490	ACCAGCCCGTC	1503	
Db	838	AACAGCTGTC	851	
RESULT 9				
LOCUS	BO720110			
DEFINITION	BO720110	904 bp	mRNA	linear
ACCESSION	AGENCOURT_8109144	lupski_sympathetic_trunk	Homo sapiens	CDNA clone
VERSION	IMAGE:6189537	5', mRNA sequence.		
KEYWORDS	BO720110.1	GI:21859007		
SOURCE	EST.	sapiens	(human)	
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 904)			
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished			
	Contact: Robert Strausberg, Ph.D.			
	Email: cgaabs-r@mail.nih.gov			
	Tissue Procurement: Dr. James R. Lupski;			
	CDNA Library Preparation: Life Technologies, Inc.			
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)			
	DNA Sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MGC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LNLN at:			
	http://image.lnl.gov			
	Plate: LLM13586	row: 0	column: 10	
	High quality sequence stop: 599.			
FEATURES				
source	location/Qualifiers			
	1..904			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:6189537"			
	/sex="male"			
	/tissue type="sympathetic trunk"			
	/dev stage="adult, 16 yr"			
	/lab_host="MDH10B"			
	/clone_lib="lupski_sympathetic_trunk"			
	/note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:			
	NotI; Site_2: SalI; CDNA made by oligo-dT priming.			
	Directionally cloned using the following adaptors:			
	5'-TCGACCAAGCGGTCG-3' and			
	5'-GACATGTTCTAGATCGCAGCGGCCGCTT(15)-3'. Size selected >			
	1 kb for average insert length 1.9 kb. This is a primary			
	library, non-amplified. Library constructed by Life			
	Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor			
	College of Medicine); available through Life			
	Technologies"			
BASE COUNT	191 a	286 c	254 g	173 t
ORIGIN				
Query Match	43.0%;	Score 740.6;	DB 13;	Length 904;
Best Local Similarity	94.7%;	Pred. No. 4.6e-159;		
Matches	810;	Conservative 0;	Mismatches 39;	Indels 6;
			Gaps 4;	
QY	243	CCGAGTCCCAACAGAGGGGCGCCCATATGGGCCCCCGGGCTCCCGGTACATGGAGCGCC	302	
Db	1	CCGAGTCCCAACAGAGGGGCGCCCATATGGGCCCCCGGGCTCCCGGTACATGGAGCGCC	60	
QY	303	GCCGTGCGACCCGGCTGGCCCCCGGGGAGATGAGCCCGCCCGAAGCGAGCGGCC	362	
Db	61	GCCGTGCGACCCGGCTGGCCCCCGGGGAGATGAGCCCGCCCGAAGCGAGCGGCC	118	
QY	363	CGCCC-GGACAGAGCGAGGACAGAGCCAGGAGCCAGCGGAGCCACCGCCCCCGCGGA	421	
Db	119	CGCCCCGAGAGCGAGGACAGAGCCAGGAGCCAGCGGAGCCACCGCCCCCGCGGA	178	
QY	422	GCCGAGTCCCAACAGAGGAAGATGCTGACAAATCTTCCTCAAGATTCGGAGC	481	

Dd		179	GCCCGAGTGGCCAAAGGAGAGATGGCTGACCAAAATTCCTCCTCAAAGATTGGAGC	238
Oy		482	TGTTCCCCGAGATCCACAGGCTTACATGAGACCTCTTGCAATTTGAGAGGAACTGGATCAAA	541
Dd		239	TGTGTCCCAGATGCCAAGCTTACTATGAGACCTCTTGCAATTTGAGAGGAACTGGATCAAA	298
Oy		542	CCATCATGCGGAGGGGGTGGACATCCAGAGGCTCTGAAGAAGGCCCATGAAGCAAAAGC	601
Dd		299	CCATCATGCGGAGGGGGTGGACATCCAGAGGCTCTGAAGAAGGCCCATGAAGCAAAAGC	358
Oy		602	GGAGGCTGCAGCTCTATATCTCCCAACCTTTAACCTGTGGAAGTCTGATGCTGAGGATTT	661
Dd		359	GGAAGCTGCAGCTCTATATCTCCCAACCTTTAACCTGTGGAAGTCTGATGCTGAGGATTT	418
Oy		662	CCGACGGCAGCATTTGCTCTGAGAGCTACGGGTGAGGGGAAAGCTCTGATGATCCCA	721
Dd		419	CCGACGGCAGCATTTGCTCTGAGAGCTACGGGTGAGGGGAAAGCTCTGATGATCCCA	478
Oy		722	GCAACAAGAGCGGAGGATCTCTCTTTCTTCAAGATTTGTCATGAGCTGGACAAG	781
Dd		479	GCAACAAGAGCGGAGGATCTCTCTTTCTTCAAGATTTGTCATGAGCTGGACAAG	538
Oy		782	ATCTTTATGCGCCCTGGAACCACTCTGCTTAGAGTGACATCGGAACCCAGACAGAGA	841
Dd		539	ATCTTTATGCGCCCTGGAACCACTCTGCTTAGAGTGACATCGGAACCCAGACAGAGA	598
Oy		842	CGAGCGGCTTCCAGAGTGAAAACGGCTGGGAGCTGATGTGCGTGCACGCTGCTCTCA	901
Dd		599	CGAGCGGCTTCCAGAGTGAAAACGGCTGGGAGCTGATGTGCGTGCACGCTGCTCTCA	658
Oy		902	TGCTGACATAACGAGCTCTCCCACTTCAACTGATATCCGCTAGCCCGGCTGCTGGGGC	961
Dd		659	TGCTGACATAACGAGCTCTCCCACTTCAACTGATATCCGCTAGCCCGGCTGCTGGGGC	718
Oy		962	TGCACACACAGAGCGGCTGAGCATTTGTCCAAGGCTCTGT-GGCAGTATGTGAAGACCAC	1020
Dd		719	TGCACACACAGAGCGGCTGAGCATTTGTCCAAGGCTCTGTGTGAAGACCAC	778
Oy		1021	AGGCTGCAGAGCTCCCATGACAAGAAATATCATCATATGAGGGG--ACAAGTATTTCCAGAGA	1078
Dd		779	AGGCTGCAGAGCTCCCATGACAAGAAATATCATCATATGAGGGGAAACATTAATTTCCACAGA	838
Oy		1079	TTTTTGAATGTGCCCC	1093
Dd		839	TTTTTGAATGTGCCCC	853
RESULT 10				
BIS53787		842 bp	mRNA	linear EST 05-SEP-2001
LOCUS		60319073.F1	NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5261913	5'
DEFINITION			mRNA sequence.	
ACCESSION		BIS53787		
VERSION		BIS53787.1	GI:15441099	
KEYWORDS		EST.		
SOURCE		Homo sapiens		
ORGANISM		Homo sapiens (human)		
REFERENCE		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE		1 (bases 1 to 842)		
JOURNAL		NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT		National Institutes of Health, Mammalian Gene Collection (MGC)		
		Unpublished		
		Contact: Robert Strausberg, Ph.D.		
		Email: ggaabs-remail.nih.gov		
		Tissue Procurement: Miklos Palkovits, M.D., Ph.D.		
		cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki		
		Toshiyuki and Piero Carninci (RIKEN)		
		cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
		DNA Sequencing by: Incyte Genomics, Inc.		
		Clone distribution: MGC clone distribution information can be		
		found through the I.M.A.G.E. Consortium/LNL at:		



```
http://image.llnl.gov
Plate: LLAM11660 row: d column: 10
High quality sequence stop: 831.
FEATURES
    source
        1..842
            Location/Qualifiers
                1..842
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5261913"
                    /tissue_type="hippocampus"
                    /lab_host="DH10B"
                    /clone_lib="NIH_MGC_95"
                    /note="Organ: brain; Vector: pBluescriptR (modified
                    pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
                    ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
                    size-selected for average insert size 2.5 kb and
                    normalized to ROT 5. This is a primary library enriched
                    for full-length clones and constructed using the
                    Cap-trapper method (Carninci, in preparation). Library
                    constructed by M. Brownstein (NIMH/NHGRI, National
                    Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      165 a   275 c   282 g   120 t
ORIGIN
Query Match      42.4%; Score 730.2; DB 12; Length 842;
Best Local Similarity 98.3%; Pred. No. 1.1e-156;
Matches 801; Conservative 0; Mismatches 8; Indels 6; Gaps 6;
QY 17 GCCGAGCCGAGCCGAGCAGGAGCGGGCGCGCGCTCCGGCCGGGGTCCCGGGG 76
DB 29 GGCCGAGCCGAGCCGAGCAGGAGCGGGCGCGCGCTCCGGCCGGGGTCCCGGGG 88
QY 77 AGCAGATCCTCAGAAATGCCCCCTTGCTGTCTGAGGCGCGGTGGCTCCGGGCCAGGCACC 136
DB 89 AGCAGATCCTCAGAAATGCCCCCTTGCTGTCTGAGGCGCGGTGGCTCCGGGCCAGGCACC 148
QY 137 GAGGGGGCACTGGATGACTCTCCAGGTGCAGGACCCCTGCCATCTATGACTCCAGGTCTTC 196
DB 149 GAGGGGGCACTGGATGACTCTCCAGGTGCAGGACCCCTGCCATCTATGACTCCAGGTCTTC 208
QY 197 AGCACCACCCACCGTGTACAGCCCGCGGGATGCCCTCTGGAGCCCGGATGCCACC 256
DB 209 AGCACCACCCACCGTGTACAGCCCGCGGGATGCCCTCTGGAGCCCGGATGCCACC 268
QY 257 A-GGGGCGCCCATGGGCCCCCGGCTCCCGTACATGGGAGCGCCCGCCCTGGAGACCCG 315
DB 269 AGGGGGCGCCCATGGGCCCCCGGCTCCCGTACATGGGAGCGCCCGCCCTGGAGACCCG 328
QY 316 GCCTGGCCCCCGGGG-ATGGAGCCCGCCCGCAAGCAGCAGCGCCCGCCCGCCCGCCG 373
DB 329 GCCTGGCCCCCGGGCATGGAGCCCGCCCGCAAGCAGCAGCGCCCGCCCGCCCGCCG 388
QY 374 GCCAGGCAACAGCCAGGGCCAGCCGAGCCAGCCCGCCCGCCCGCCCGCCCGCCGAGTGCCA 433
DB 389 GCCAGGCAACAGCCAGGGCCAGCCGAGTGTCCCGCCCGCCCGCCCGCCCGCCGAGTGCCA 448
QY 434 AGAGGAGGAAGATGCTCACAAAATCTCTCCCTCAAGGATTCGGGAGCTGCTCCCGAGT 493
DB 449 AGAGGAGGAAGATGCTCACAAAATCTCTCCCTCAAGGATTCGGGAGCTGCTCCCGAGT 508
QY 494 CCCAGGCTTACATGGACCTCTTGGCAATTTGAGAGGAAACTGGATCAAAACCATCATGCGGA 553
DB 509 CCCAGGCTTACATGGACCTCTTGGCAATTTGAGAGGAAACTGGATCAAAACCATCATGCGGA 568
QY 554 AGGGGTGGACATCCAGAGGCTCTGAAGAGGCCCATGAAGCAAAAGCGGAGCTGGAC 613
DB 569 AGCGGTGGACATCCAGAGGCTCTGAAGAGGCCCATGAAGCAAAAGCGGAGCTGGAC 628
QY 614 TCTATATCTCAACACTTTTAACTCCTCGGAAGTCTGATGCTGAGGATTTCCGACGGCAGCA 673
DB 629 TCTATATCTCAACACTTTTAACTCCTCGGAAGCTGATGCTGAGGATTTCCGACGGCAGCA 688
QY 674 TTGCTCTCTGGAGCTACCGGTGGAGGGAAGCTCTCTGGATGATCCCGAGCAAAACAGAGC 733
```

```
|||||
689 TTGCTCTCTGGAGCTACCGGTGGAGGGAAGCTCTCTGGATGATCCCGAGC-AAACAGAGC 747
QY 734 GGAAGTTCTCTTTCTTCAAGAGTTTGTCTATCGAGCTGGACAAAGATCTTTTATGGCC 793
DB 748 GGAAGTTCTCTTTCTTCTTCAAGAG-TTGTCTATCGAGCTGGACAAAGATCTTTATGGCC 806
QY 794 CTGACACCACTCTGTTGAGTGGCATCGGACACC 828
DB 807 CTGAC-AGCACCCTGTTGAGTGGCATCGGAAACC 840

RESULT 11
BI520621/c
LOCUS
DEFINITION
603071572T1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5163880 3',
mRNA sequence.
ACCESSION
BI520621
VERSION
BI520621.1 GI:15345413
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 807)
NIH-MGC http://mgc.nci.nih.gov/.
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)
UNPUBLISHED
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11406 row: o column: 17
High quality sequence start: 29
High quality sequence stop: 805.
FEATURES
    source
        1..807
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5163880"
            /tissue_type="medulla"
            /lab_host="DH10B"
            /clone_lib="NIH_MGC_119"
            /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
            Site 2: EcoRV (destroyed); RNA source normal medulla from
            anonymous male age 27. Library is oligo-dT primed and
            directionally cloned (EcoRV site is destroyed upon
            cloning). Average insert size 1.3 kb, insert size range
            0.9-3 kb. Library is normalized and enriched for
            full-length clones and was constructed by C. Gruber
            (Invitrogen). Research Genetics tracking code 013. Note:
            this is a NIH_MGC Library."
BASE COUNT      150 a   219 c   242 g   196 t
ORIGIN
Query Match      42.3%; Score 728.8; DB 12; Length 807;
Best Local Similarity 96.3%; Pred. No. 2.2e-156;
Matches 778; Conservative 0; Mismatches 27; Indels 3; Gaps 3;
QY 867 TGGGACCTGAGTGGCTGCGCTGCGCTGCTCTCATGCTGAGCTACAGAGCTCCCGAGTT 926
DB 807 TGGGACCGGAGTG-GCGCTGCGCTGCTCTCATGCTGAGCTACAGAGCTCCCGAGTT 749
QY 927 CAAACTGATCCCCCGCTAGCCGCTGCTGGGCTGCACACAGAGCCGCTCAGCCAT 986
DB 748 CAAACTGATCCCCCGCTAGCCGCTGCTGGGCTGCACACAGAGCCGCTCAGCCAT 689;
```





```
Db 611 CTCCTGTCCACGGTCAACAGCAGAGATCATGTCTCTGGACAGTAAGATCCATGAGACG 670
Qy 1324 ATTGAGTCCATAAACACAGCTCAAGATCCAGAGGAGCTTCATGCTAAGCTTCTCCAGAGAC 1383
Db 671 ATTGAGTCCATAAACACAGCTCAAGATCCAGAGGAGCTTCATGCTAAGCTTCTCCAGAGAC 730
Qy 1384 CCCAAAGGCTATGTCCAAAGACCTGCTCCGCTCCCAAGAGCCGGAGCCTCAAGGTG-ATGAC 1442
Db 731 CCCAAAGGCTAGTCCAAAGACCTGCTCCGCTCCCAAGAGCCGGATCTCAAGGTGAATGAC 790
Qy 1443 AGATGTAGCCGGCAACCTGAAGAGGAGCGC-CGGGCTGAGTTCTACCAACCA-GCCCTGG 1500
Db 791 CGATGTGCAGGGAACCCCGAGGAAGACGCGAGGGGCTGAGTTCTACCAACAGGTCTCTGG 850
Qy 1501 TCCAGGAGGCGGTGAG-TCCGCTACTTCTGCAAGATCCAGAGCCG-AGGCAGGAGC 1558
Db 851 TCCAGGAGGCGGTGAGCCGCTACTTCTGATGATCCAGAGCCGCAAGGCGAGGAGC 910
Qy 1559 TGGAGCAGTCCGTGGTT-GTGGCAACACCTAGGAGCCC 1596
Db 911 TGAAGCAGTCCGTGGTTCGTCGCAACAGCTAGGAGCCC 949

RESULT 13
LOCUS AL533695 1201 bp mRNA linear EST 12-MAY-2003
DEFINITION AL533695 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF002YC19 5-PRIME, mRNA sequence.
ACCESSION AL533695
VERSION AL533695.2 GI:30538791
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 13, 2001 this sequence version replaced gi:12797188.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: secre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4359.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF002AB10Qp1&cluster=4359.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF002AB10Qp1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF002YC19"
/tissue_type="FETAL BRAIN"
/dev_stages="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 274 a 332 c 353 g 198 t 44 others
ORIGIN

Query Match 40.5%; Score 698.4; DB 9; Length 1201;
Best Local Similarity 96.0%; Pred. No. 2.4e-149;
Matches 789; Conservative 19; Mismatches 5; Indels 9; Gaps 9;
```

```
Qy 219 GCGCCCCGGATGCGTCTTGAGAGCCGGATGCCCCACCA-GGGCGGCCCATGGGCCCCCC 277
Db 226 GCGCCCCGGATGCGTCTTGAGAGCCGGATGCCCCACCAAGGGGGCGCCCATGGGCCCCCC 285
Qy 278 GGGTCTCCCGTACATGAGGAGAGCCCGCGCTGCGACCCGGCTGCGCCCGCGGGG-ATGGA 336
Db 286 GGGTCTCCCGTACATGAGGAGAGCCCGCGCTGCGACCCGGCTGCGCCCGCGGCATGGA 345
Qy 337 GCGCGCCGGAAGCGAGCAGCGCCCGCGCC-GGCAGAGCCAGGCACAGAGCCAGGGCCA 395
Db 346 GCGCGCCGGAAGCGAGCAGCGCCCGCGCCGGGAGAGCCAGGCACAGAGCCAGGGCCA 405
Qy 396 GCGGAGCCCAACCGCCCGCGCGGAGCCGAGTGCACAGAGGAGGAAGATGGCTGACAA 455
Db 406 GCGGAGCCCAACCGCCCGCGCGGAGCCGAGTGCACAGAGGAGGAAGATGGCTGAMAA 464
Qy 456 AATCCTCCCTCAAAAGGATTCGGGAGCTGGTCCCGAGTCCAGGCTTACATGAGACCTCTT 515
Db 465 AATCCTCCCTCAAAAGGATTCGGGAGCTGGTCCCGAGTCCAGGCTTACATGAGACCTCTT 524
Qy 516 GGCATTTGAGAGGAAACTGGATCAAAACCATCATCGGAAGCGGGTGGACATCCAGGAGGC 575
Db 525 GGCATTTGAGAGGAAACTGGATCAAAACCATCATCGGAAGCGGGTGGACATCCAGGAGGC 584
Qy 576 TCTGAAGAGGCCCATGAAGCAAAAGCGGAAGCTCGGACTCTATATCTCTCAACACTTTTAA 635
Db 585 TCTGAAGAGGCCCATGAAGCAAAAGCGGAA-STGCGACTCTATATCTCTCAACACTTTTAA 643
Qy 636 CCCTGCGAAGTCTGATGCTGAGGATTCGAGCGGAGCATTTGCCCTCTGGGAGCTACGGGT 695
Db 644 CCCTGCGAA-CCTGATGCTGAGGATTCGAGCGGAGCATTTGCCCTCTGGGAGCTACGGGT 702
Qy 696 GGAGGGGAAGCTCCTGATGATCCAGCAAAACAGAGCGGAAGTCTCTCTCTTCTTCAA 755
Db 703 GGAGGGGAAGCTCCTGATGATCCAGCAAAACAGAGCGGAAGTCTCTCTCTTCTTCAA 762
Qy 756 GAGTTTGTGTCATCGAGTGGCAAAAGATCTTTATGGCCCTGACAAACCACTCGTTGAGTG 815
Db 763 GAGTTTGTGTCATCGAGTGGCAAAAGATCTTTATGGCCCTGACAAACCACTCGTTGAGTG 822
Qy 816 GCATCGACACCCACGACCCAGGAGCGGCTTCCAGGTGAAACGGCTT-GGGAGC 874
Db 823 GCATCGACACCCACGACCCAGGAGCGGCTTCCAGGTGAAACGGCTTGGGGAGC 882
Qy 875 TGAGTGGCGTGCACCGCTCTCTCATGCTGGACTACGAGCTCCCGAGTTCAACTGG 934
Db 883 TGAGTGGCGTGCACCGCTCTCTCATGCTGGACTACGAGCTCCCGAGTTCAACTGG 942
Qy 935 ATCCCGCCCTAGCCCGCTGCTGGGGCTGCACACAGAGCGGCTCAGCCATTGTCAGG 994
Db 943 AT-CCCGCCCTAGCCCGCTGCTGGGGTGCACACACAGAGCGGCTCAGCCATTGTCAGG 1001
Qy 995 CCCTGTGGCAGTATGTGAAGACCAACAGGCTGAGGAGCTCCC 1036
Db 1002 CCCTGTGGCAGTATGTGAAGACCAACAGGCTGAGGAGCTCCC 1042

RESULT 14
LOCUS CD517671 837 bp mRNA linear EST 06-JUN-2003
DEFINITION AGENCOURT 14372899 NIH MGC 181 Homo sapiens cDNA clone
IMAGE:30400116 5', mRNA sequence.
ACCESSION CD517671
VERSION CD517671.1 GI:31449389
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 837)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
```

## JOURNAL COMMENT

Unpublished  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cga@bms-r@mail.nih.gov  
Tissue Procurement: Dr. Michael Brownstein  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM487 row: 1 column: 13  
High quality sequence stop: 614.  
Location/Qualifiers

## FEATURES

source

1. 837  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30400116"  
/tissue\_type="White Matter"  
/dev\_stage="Unknown"  
/lab\_host="DH10B-Ton A (T1 and T5 phage resistances)"  
/clone\_lib="NIH MGC 181"  
/note="Vector: PCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV  
(destroyed); Library is oligo-dT primed and directionally  
cloned (EcoRV site is destroyed upon cloning). Average  
insert size 1.42 kb. Library was constructed by  
(Invitrogen). Note: this is a NIH\_MGC Library."

## BASE COUNT

174 a 276 c 258 g 129 t

## ORIGIN

Query Match 40.3%; Score 694.4; DB 14; Length 837;  
Best Local Similarity 96.8%; Pred. No. 1.7e-148;  
Matches 784; Conservative 0; Mismatches 16; Indels 10; Gaps 7;

218 AGCGCCCGGAGTCCGCTGAGAGCCCGGATGCCACCA-GGGGGGCCCATGGGCCCC 276  
|||||  
1 AGCGCCCGGAGTCCGCTGAGAGCCCGGATGCCACCA-GGGGGGCCCATGGGCCCC 60  
277 CGGGCTCCCGTACATGGGAGCCCGCGTGGAGCCCGGCTGGCCCGCGCGG-ATGG 335  
61 CGGGCTCCCGTACATGGGAGCCCGCGTGGAGCCCGGCTGGCCCGCGCGG-ATGG 120  
336 AGCGCCCGGAGTCCGCTGAGAGCCCGGATGCCACCA-GGGGGGCCCATGGGCCCC 394  
121 AGCGCCCGGAGTCCGCTGAGAGCCCGGATGCCACCA-GGGGGGCCCATGGGCCCC 180  
395 AGCGCCCGGAGTCCGCTGAGAGCCCGGATGCCACCA-GGGGGGCCCATGGGCCCC 454  
181 AGCGCCCGGAGTCCGCTGAGAGCCCGGATGCCACCA-GGGGGGCCCATGGGCCCC 240  
455 AATCTCTCCCTCAAGAGATTCGGAGCTGGTCCCGAGTCCCGAGGCTTACATGACCTCT 514  
241 AATCTCTCCCTCAAGAGATTCGGAGCTGGTCCCGAGTCCCGAGGCTTACATGACCTCT 300  
515 TGGCATTTGAGAGAACTGATCAACATCATGCGGAAGCGGGTGAATCAGAGAG 574  
301 TGGCATTTGAGAGAACTGATCAACATCATGCGGAAGCGGGTGAATCAGAGAG 360  
575 CTCTGAAGAGCCCAAGAAAGCGAAAGTGGAGCTCTATATCTCCAACTTTTA 634  
361 CTCTGAAGAGCCCAAGAAAGCGAAAGTGGAGCTCTATATCTCCAACTTTTA 420  
635 ACCCTGGAAGTGTGATGCTGAGATTCGAGCGGAGCATTTGCTCTGGAGCTACGG 694  
421 ACCCTGGAAGTGTGATGCTGAGATTCGAGCGGAGCATTTGCTCTGGAGCTACGG 480  
695 TGGAGGGGAAGCTCTGATGATCCAGCAACAGAGGGAAGTTCTCTTTCTTCA 754  
481 TGGAGGGGAAGCTCTGATGATCCAGCAACAGAGGGAAGTTCTCTTTCTTCA 540

QY 755 AGAGTTGGTATCGAGCTGAGCAAAAGTCTTTATGGCCCTGACCAACCACTGTTAGT 814  
|||||  
DB 541 AGAGTTGGTATCGAGCTGAGCAAAAGTCTTTATGGCCCTGACCAACCACTGTTAGT 600  
QY 815 GGCATCGACACCCACGACCCAGAGAGCGAGCGGCTTCCAGGTAAAGCGCTT-GGGAGC 873  
601 GGCATCGACACCCACGACCCAGAGAGCGAGCGGCTTCCAGGTAAAGCGCTTGGGAGC 660  
QY 874 CTGAGTGGCGGTGAGAGCGTGTCTCTCATGCTGAGCTACACACCTCCCAAGTTCAAAC 933  
661 CTGAGTGGCGGTGAGAGCGTGTCTCTCATGCTGAGCTACACACCTCCCAAGTTCAAAC 720  
QY 934 GATCCCGGCTGAGCGCGG--CTGCTGGGCTGACACACAGAGCGG--CTCAGCCATTGT 989  
721 GATCCCGGCTGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
DB 990 CCAGGCCC-TGTGCGATGATGTAAGAC 1017  
781 CCAGGCCCCTGTGCGATGATGTAAGAC 810

RESULT 15  
CD558853  
LOCUS CD558853 894 bp mRNA linear EST 11-JUN-2003  
DEFINITION AGENCOURT 14394871 NIH MGC 181 Homo sapiens CDNA clone  
IMAGE:30400116 5', mRNA sequence.  
ACCESSION CD558853  
VERSION CD558853.1 GI:31584921  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 894)  
AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished

## COMMENT

Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics / NIH  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: cga@bms-r@mail.nih.gov  
Tissue Procurement: Dr. Michael Brownstein  
CDNA Library Preparation: Invitrogen Corp  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: NDAM487 row: 1 column: 13  
High quality sequence stop: 615.  
Location/Qualifiers

## FEATURES

source

1. 894  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:30400116"  
/tissue\_type="White Matter"  
/dev\_stage="Unknown"  
/lab\_host="DH10B-Ton A (T1 and T5 phage resistances)"  
/clone\_lib="NIH MGC 181"  
/note="Vector: PCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV  
(destroyed); Library is oligo-dT primed and directionally  
cloned (EcoRV site is destroyed upon cloning). Average  
insert size 1.42 kb. Library was constructed by  
(Invitrogen). Note: this is a NIH\_MGC Library."

## BASE COUNT

195 a 289 c 270 g 139 t

## ORIGIN

Query Match 39.7%; Score 685.2; DB 14; Length 894;  
Best Local Similarity 96.2%; Pred. No. 2.3e-146;  
Matches 744; Conservative 0; Mismatches 24; Indels 5; Gaps 4;

